22

12-

58

Psychological Bulletin

EDITED BY

LYLE H. LANIER, NEW YORK UNIVERSITY

WITH THE CO-OPERATION OF

S. H. BRITT, McCANN-ERICKSON, INC., NEW YORK; D. A. GRANT, UNIVERSITY OF WISCONSIN; W. T. HERON, UNIVERSITY OF MINNESOTA; W. A. HUNT, NORTHWESTERN UNIVERSITY; D. G. MARQUIS, UNIVERSITY OF MICHIGAN; A. W. MELTON, OHIO STATE UNIVERSITY; J. T. METCALF, UNIVERSITY OF VERMONT.

VOLUME 46, 1949

PUBLISHED BI-MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W.,
WASHINGTON 5, D. C.

Bureau Ednl. Psy. Rese

12.8.70
J 146

CONTENTS OF VOLUME 46

ORIGINAL CONTRIBUTIONS, GENERAL REVIEWS, DISCUSSIONS

Infant care and personality: Harold Orlansky 1
A note on Grant's discussion of the Latin square principle in the design and analysis of psychological experiments: B. R. Bugelski 49
Note on the Postman-Stone controversy: Jack Bernard 51
The Prevost-Fechner-Benham subjective colors: Jozef Cohen and Donald A. Gordon 97
An extension of control group design: Richard L. Solomon 137
Animal hypnosis: Thelma Gilman and F. L. Marcuse 151
Problems of psychometric scatter analysis: Joseph Jastak 177
Numerical transformations in the analysis of experimental data: C. G. Mueller 198
Historical beginnings of child psychology: Wayne Dennis 224
A historical survey of psychology in the Soviet Union: Ivan D. London 241
Statistical bias in comparisons of monaural and binaural thresholds: Binaural summation or binaural supplementation: Moncrieff H. Smith and J. C. R. Licklider 278
The genetics of schizophrenia: Nicholas Pastore 285
Notes on "The validity of personality inventories in military practice" by Ellis and Conrad: Doncaster G. Humm and Kathryn A. Humm 303
Reply to the Humms' "Notes on the 'Validity of personality inventories in military practice' ": Herbert S. Conrad and Albert Ellis 307
Stimulus generalization of conditioned responses: Gregory Razran 337
Psychotherapy as a problem in learning theory: Edward J. Shoben 366
Statistical methods applied to Rorschach scores, A review: Lee J. Cronbach 393
The use and misuse of the chi-square test: Don Lewis and C. J. Burke 433
Taste blindness to phenyl-thio-carbamide and related compounds: Jozef Cohen and Donald P. Ogdon 490
Army Air Forces Aviation Psychology Program Research Reports: Special Reviews 499

BOOK REVIEWS

Boring, Langfeld and Weld's Foundations of psychology: WALTER S. HUNTER 54
Dennis' Readings in the history of psychology: EDWIN G. BORING 57
Lewis' Quantitative methods in psychology: C. H. GRAHAM AND C. G. MUELLER 63
Ruch's Psychology and life: DONALD E. LUNDBERG 66
Davis' Educational psychology: FREDERICK B. DAVIS 67
Ghiselli and Brown's Personnel and industrial psychology: JOHN P. FOLEY, JR. 68
Burtt's Applied psychology: S. RAINS WALLACE 70
Otis and Leukart's Job evaluation: C. H. LAWSHE 73
Kaplen's Encyclopedia of vocational guidance: STANLEY G. DULSKY 74
Naumburg's Studies of "free" art expression of behavior problem children and adolescents as a means of diagnosis and therapy: MARIA A. RICKERS-OVSIANKINA 76
Lindner and Seliger's Handbook of correctional psychology: R. J. CORSINI 78
Cottrell and Eberhart's American opinion on world affairs in the atomic age: STEUART HENDERSON BRITT 79
Wolberg's Medical hypnosis: F. L. MARCUSE 81
Hall's Annual review of physiology: BERNARD F. RIESS 82
Moore's The driving forces of human nature and their adjustment: GEORGE W. HARTMANN 83
Poffenberger's James McKeen Cattell: man of science. 1860–1944: MARK A. MAY 85
Gorer's The American people: PERCY BLACK 89
Evans' Introduction to color: HARRY HELSON 166
Doob's Public opinion and propaganda: JEROME S. BRUNER 169
Sherif's An outline of social psychology: GLEN HEATHERS 172
Halstead's Brain and intelligence: DONALD B. LINSDLEY 173
Wiener's Cybernetics: FRANKLIN V. TAYLOR 236
Sechenov's Selected works: EDWIN G. BORING 309
Blum's Industrial psychology and its social foundations: WILLARD A. KERR 311
Hollingworth's Psychology and ethics: GEORGE W. HARTMANN .. 313
Vaughan's Social psychology: RICHARD A. LITTMAN 314
Peterson, Marzolf and Bayley's Educational psychology: GLENN M. BLAIR 317

Bray's Psychology and military proficiency: Glen Finch 318
Eysenck's Dimensions of personality: Leona E. Tyler 319
White's The abnormal personality: T. W. Richards 321
MacKinnon and Henle's Experimental studies in psychodynamics: Howard H. Kendler 322
Ewen's Mental health: Ludwig Immergluck 323
Mendel's Personality in handwriting: Werner Wolff 324
Gesell and Amatruda's Developmental diagnosis: Margaret R. Kuenne 325
Tyler's The psychology of human differences: Robert L. Thorndike 326
Alschuler and Hattwick's Painting and personality: Lawrence Joseph Stone 327
Pennington and Berg's An introduction to clinical psychology: Ralph D. Norman 330
Hsu's Under the ancestor's shadow: John W. Gardner 430
Flanagan's The aviation psychology program in the Army Air Forces: Lyle H. Lanier 499
DuBois: The classification program: R. M. Bellows 501
Thorndike's Research problems and techniques: A. S. Thompson 503
Melton's Apparatus tests: R. H. Seashore 504
Guilford's Printed classification tests: T. W. Harrell 506
Davis' The AAF qualifying examination: H. S. Conrad 508
Gibson's Motion picture testing and research: W. H. C. Prentice 511
Miller's Psychological research on pilot training: J. L. Finan ... 512
Kemp and Johnson's Psychological research on bombardier training: D. D. Wickens 514
Carter's Psychological research on navigator training: J. L. Kennedy 515
Hobbs' Psychological research on flexible gunnery training: W. E. Kappauf, Jr. 516
Cook's Psychological research on radar observer training: C. S. Bridgman 518
Dailey's Psychological research on flight engineer training: G. R. Stone 519
Wickert's Psychological research on problems of redistribution: D. E. Baier 521
Bijou's Psychological program in AAF convalescent hospitals: G. A. Kelly 523
Crawford's Psychological research on operational training in the continental air forces: Carl H. Wedell 524

Lepley's Psychological research in the theaters of war: N. H. PRONKO 525
Deemer's Records, analysis and test procedures: E. R. HENRY 526
Fitts' Psychological research on equipment design: L. C. MEAD 527
Hilgard's Theories of learning: NEAL E. MILLER 529
Woodworth's Contemporary schools of psychology: HORACE B. ENGLISH 532

MISCELLANEOUS

Books and Materials Received 94, 238, 334, 431
Index of Subjects 535
Index of Authors 537

138H

Vol. 46, No. 1 January, 1949

Psychological Bulletin

INFANT CARE AND PERSONALITY[1]

HAROLD ORLANSKY
Department of Anthropology, Yale University

INTRODUCTION

The importance of childhood experience to the formation of personality was one of the great findings of psychoanalysis. No social scientist will dispute that general finding, but all will wish to elucidate it.

The orthodox school of psychoanalysis, in keeping with Freud's genetic and biological orientation, has stressed the special significance of infantile suckling, excretory, and genital functions for the shaping of adult personality. Indeed, the Freudian oral-anal-genital characterology rests upon the assumption that specific channeling of these infantile physiological urges by varied parental care elicits specific psychological constellations in the individual.[2] The essential features of this characterology have been borrowed by some psychologists (**116**, pp. 362–385). Reasoning from Freudian theory, pediatricians and psychiatrists have advocated systems of infant care which they believe will promote the growth of secure and unneurotic personalities.[3] In general, these systems advise prolonged breast feeding on self-demand by the infant, gradual and late weaning, and the late and lenient induction of bowel and bladder training. From the same theoretical base, some anthropologists and psychoanalysts have, in recent years, attempted to explain presumed character differences between members of different cultures in terms of their different patterns of infant discipline.[4]

---

[1] The writer wishes to thank Drs. Louise Ames, Wendell Bennett, Ralph Linton, and Robert Sears for having been kind enough to read and criticize this paper in manuscript. If he had the ability and energy to utilize all their advice, it would be a better paper.

[2] For statements of this characterology, see Freud (**43**; **44**; **46**, pp. 135–140), Abraham (**1**, pp. 248–279, 370–417), Glover (**54**), Jones (**77**), Lorand (**96**), and Fenichel (**36**, pp. 62–69, 278–284, 487–496).

[3] A very extensive bibliography could be cited here which would be swollen by each succeeding issue of such journals as *Psychiatry* and the *American Journal of Orthopsychiatry.* Frank (**38, 39, 40**), Ribble (**128**), and Moloney (**107, 108, 109**) are among the ardent advocates of this program.

Freudian theory, however, arose from clinical investigations of adult character, and the reconstructions of infant experience obtained from the analytic couch do not constitute empirical findings on the infantile situation. The work of child analysts like Melanie Klein and Anna Freud has added something to our direct knowledge of child psychology, but they have generally been more interested in applying psychoanalytic theory than in testing it. After a comprehensive review of the evidence, it was thus possible for three psychologists to say that "although we have now been exposed for some time to psychoanalytic and other psychiatric hypotheses regarding the effects of birth trauma, weaning trauma, extreme emphasis on early control of urination and defecation . . . we have almost no objective records of the development of children going through these experiences, or of experiments controlling certain aspects of the problem" (**113**, p. 575). Ten years later, Murphy opined that "some aspects" of these hypotheses had been experimentally confirmed, but "the observations are meager and the theory overpretentious" (**114**, p. 485).

Sears (**136**) has broadly surveyed objective studies designed to test psychoanalytic theory, but no one has yet concentrated attention upon the subject of infant disciplines and compiled the factual evidence on the relation between these disciplines and the development of personality. The present paper proposes to do this, gathering data from two sources: (a) empirical, experimental, and semi-experimental investigations in Western culture, and (b) evidence obtained from anthropological inquiries into other cultures. The writer is under no illusion as to the adequacy of much of the data which will be presented, and criticism will be offered where it seems appropriate. Nevertheless, he is of the opinion that the evidence permits definite, if negative, conclusions in regard to the effect of certain features of infant care upon personality; and where the evidence is less satisfactory, it is to be hoped that others will be encouraged to seek an empirical resolution of the problem.

[4] The writings of Gorer (**58, 59**), Erikson (**30, 31, 32**), LaBarre (**84**), and Roheim (**134**) illustrate the simplicist assertion that infant disciplines determine group character and culture. Their theoretical position should be distinguished from the efforts of such students as Kardiner (**79, 80**) and Du Bois (**26, 27**) to explain group character in terms of social and economic institutions and broad cultural patterns.

FEATURES OF INFANT CARE

Nursing Experiences

It has been argued that many features of the infant's early feeding experiences contribute to the formation of personality. Among the factors to which special psychological significance has been attributed are breast vs. artificial feeding, length of breast feeding, self-demand vs. scheduled feeding, and weaning. Studies which have attempted to test the significance of these factors are summarized below, followed by a brief evaluation of their validity.

Breast vs. bottle feeding. The chairman of the Committee on Maternal and Child Feeding of the National Research Council concluded a recent investigation of the advisability of breast vs. artificial feeding with the following statement:

> Although there is a voluminous literature on the subject of the emotional value of breast feeding to both mother and baby, it is regrettable that concrete evidence on this point is difficult to obtain. This is particularly regrettable in view of the rather emotional treatment the subject is receiving in both professional and lay circles. The necessary evidence will have to come from long observations on the development of the personalities of both breast-fed and artificially fed infants (**3**, p. 915).

While really adequate evidence is, to be sure, unavailable, the following studies have come closest to approximating the required observations.

Rogerson and Rogerson (**131**) traced 109 children who had been observed as infants in their London clinic seven years before. Those artificially fed during infancy had experienced more feeding difficulties, poorer health, and greater restlessness during infancy than had those breast fed during infancy. At the age of 7, the group which had been artificially fed in infancy was found to have a greater proportion of its members manifest poor appetite, poor sleep, enuresis, fears, nervousness, and a poorer showing in school than the breast-fed group. The authors conclude that "there is evidence strongly suggesting that there is a positive association between feeding difficulties in infancy and subsequent psychological difficulties at elementary school age" (p. 1181).

In a study at Joliet, Ill., Hoefer and Hardy (**67**) subjected 383 children 7–13 years old to physical and mental tests, correlating their ratings on the tests with the length of breast feeding in infancy. These findings will be reviewed below, but at this point we are interested in their observation that "Children who were artificially fed were, on the whole, inferior physically and mentally to the breast fed" (**67**, p. 619). The artificially fed children were more poorly nourished, more susceptible to diseases, and slower in learning to talk and to walk than were the breast-fed children.

Both the Rogerson and Rogerson and the Hoefer investigations were based on general populations, with no effort having been made to equate the groups studied for factors other than method of feeding in infancy. The physical and psychological difficulties which they found among artificially fed children can be accepted as descriptively accurate for their populations; but the inference that artificial feeding was a *causative element* in the creation of the difficulties is not warranted by these data. Indeed, Rogerson and Rogerson expressly disavow the propriety of inferring such a causal relationship (**131**, p. 1173). That ill-health and feeding difficulties in infancy can often be taken as evidence of emotional maladjustment is, of course, widely recognized.

Our inference that these difficulties were not necessarily due to the manner of infant feeding is supported by a study conducted by Faber and Sutton (**34**) which comes closer to the scientific ideal of controlled experiment than do the studies outlined above. These authors compared 42 babies at the Stanford University clinic who had been breast-fed less than six weeks and were thereafter reared on bottled milk with an equal number of babies who had been breast-fed for an average of forty weeks. In other respects than the type of milk, the diets of both groups were alike; 26 members of each group were paired for weight at birth, and all babies received "excellent care." It was found that the breast-fed babies gained more weight than the bottle-fed babies during the first quarter year, but this relationship was reversed during the rest of the year. The bottle-fed group had better health and fewer diseases than the breast-fed group. Accepting the psychosomatic criterion that physical and psychological health are apt to be correlated, this study would seem to show that artificial feeding need be in no way inferior to breast feeding in promoting the emotional well-being of the infant.

Summarizing the picture in the light of present data, it does not appear to be possible to establish preferential norms of breast or bottle feeding in our society without consideration of the class position of the parents and the psychological condition of the mother. The norm for lower-class families in America and England apparently entails breast feeding of children in a larger proportion of cases than does the norm for middle-class families;[5] different hospitals in the same city regularly have larger or smaller proportions of breast-fed babies in accordance (in part) with the class difference of their clientele (**13**). There is reason to believe that bottle feeding introduced into the average lower-class home might often have beneficial effects upon the child:

> Babies between birth and about five months of age, when not breast-fed, . . . develop better in our nursery than in the average proletarian household.

---

[5] See Ericson (**29**), Rogerson and Rogerson (**131**, p. 1166), and Davis and Havighurst (**18**).

> Their gain in weight is more regular and intestinal disturbances are less frequent. . . . In times of illness, the absence of tension and anxiety which the young mother invariably feels is certainly of advantage to the child.[6]

Above all, the character of the mother must be evaluated before an intelligent opinion can be ventured as to whether breast or artificial feeding will better serve to promote the healthy development of the child's personality:

> In some cases, even though supplementary feeding may be necessary, breast feeding can advantageously be continued to foster the mother-child relationship. In other cases, it may be advisable, in order to minimize maternal rejection, to reduce breast feedings, despite the physical possibility of continuing them (**48**, p. 235).

We have seen no evidence to favor the current belief among some pediatricians that breast feeding is inevitably more advantageous to the child than artificial feeding.[7]

Length of breast feeding. In the study described above, Hoefer and Hardy found that the highest ratings in muscle tone, physical and mental development, and early age at walking and dentition were registered by children who, as infants, had been breast-fed 4–9 months; children breast-fed 3 months or less and 10 months or more had lower ratings than did this medial group (**67**).

Somewhat comparable results were obtained by Pearson in a study of 72 problem children at the Philadelphia Child Guidance Clinic. Children breast-fed less than 6 months or more than 10 months were more maladjusted at school and displayed timidity and nervousness with greater frequency than did those children who had been breast-fed from 6–10 months (**117**, p. 289).

Childers and Hamil (**15**) examined the records of 469 problem children at the Harper Hospital clinic in Detroit, and correlated the fre-

[6] Freud and Burlingham (**42**, p. 12). It is only fair to report that the authors add: "Breast-fed babies are, of course, better off than bottle-fed babies wherever they are. Our best results are found in babies who are breast-fed by their own mothers in our home" (p. 13). There are obvious reasons why a nursery baby who receives individual care from its mother should progress better than one who shares with many other babies the services of a professional nurse; but the assertion that breast-fed babies are invariably better off than bottle-fed babies remains unproven.

[7] Since breast-feeding is so frequently considered *de rigeur*, it may be well to note that a contrary opinion has been expressed by two outstanding pediatricians: "[Some people] . . . deplore the emotional losses which the bottle-fed baby suffers. It is probable that the nervous system of the modern infant is adaptive enough to meet most of the demands made by new modes of nutrition. Indeed, oncoming generations of infants by virtue of these very changes may even acquire emotional patterns of self-dependence which will ultimately add to the mental stamina of the race and form personalities better suited to the stresses of civilization" (Gesell and Ilg, **52**, p. 42).

quency of occurrence of 95 "personality traits" with the duration of breast feeding in infancy. Children breast-fed 1–5 months showed a higher percentage of "abnormal behavior manifestations" than children breast-fed less than 1 month or more than 5 months. However, this study is quite unsatisfactory methodologically, and its results can be given little credence.

Maslow and Szilagyi-Kessler (**102**) gave a security-insecurity inventory to students in classes at Brooklyin College and correlated this measure of security with the length of infantile breast feeding. "The highest security scores were obtained by the groups who were breast-fed little or not at all and by those who were fed at the breast for over a year" (**102**, p. 84).

Peterson and Spano (**118**) correlated the length of breast feeding with personality ratings of nursery school children. They found "There is no evidence that those children breast fed least are more maladjusted and insecure than those breast fed longer"; and, again, "in the group of normal children studied . . . there is no relation between duration of breast feeding and child personality at the nursery school level, nor with adjustment at the pre-adolescent stage" (**118**, pp. 63, 65).

All five investigations, it should be noted, agree that there is no linear correlation to be observed between length of breast feeding and any major aspect of personality studied. This is the only definitive conclusion that present evidence seems to permit. On the basis of the Hoefer and Hardy, and Pearson studies one might speculate that a medial duration of breast feeding is more apt to be associated with "normal" personality development in our society than is either a brief or prolonged period of breast feeding, but this hypothesis is flatly contradicted by the Maslow and Szilagyi-Kessler finding. Barring questions as to the validity of the personality measures employed, it is clear that extreme caution is necessary in making any generalization applying to large populations from the very inadequate and unrepresentative samples upon which almost all of these studies were based. This is a conclusion to which we will have to return later.

The danger of such generalization can be illustrated by Levy's statement that "In general, all factors favoring rejection of the child tend to shorten, all factors favoring overprotection tend to lengthen, the breast-feeding act" (**91**, p. 59; cf. also **90**, p. 101). Levy has probably contributed more to the experimental solution of the problems with which we are concerned than any other person, and his procedures are models in the field. The statement quoted above is based primarily upon research by Freeman (**41**) in which 100 long (12 months or over) and 100 short (1 month or less) breast-feeding cases were selected from a group of children at the New York Institute for Child Guidance, using no other criterion. It was then found that half of the children in the short-fed group were "rejected" and two-thirds of the children in the long-fed group were "over-protected' by their mothers. There is no reason to dispute the accuracy of this finding or the reasonableness of Levy's generalization. But Peterson and Spano tested the generalization in their study of nursery school children, and concluded that "duration of breast feeding is not related to maternal rejection and cannot therefore be taken as an index of rejection"(**118**, p. 65). There may, of course, be some error in their procedure, although this is not apparent from a reading of the report. But is it just as likely that the groups studied differed in certain essential cultural characteristics, and that the contrasting generalizations advanced by Levy and Peterson are each valid only for populations which resemble the original groups in these (unexplored) characteristics.

Self-demand vs. scheduled feeding. The prevalent opinion in many circles is that self-demand feeding in which nursing always rewards the child's crying is better calculated to produce a confident and unneurotic personality than is rigidly scheduled feeding in which the clock rather than the infant sets the nursing time. The following quotations illustrate this viewpoint:

> To the white child, whose feeding and other routines are rigidly scheduled, the mother or nurse . . . must appear incalculable. . . . He must develop a feeling that each individual is alone in life.
>
> To the Navaho baby, on the other hand, other persons must appear warmer and more dependable, for every time he cries, something is done for him (**85**, p. 31; **82**, p. 57).

> The parent who responds to a clock rather than to the behavior of the child is, from the child's point of view, not responsive at all. . . . If the rewards which a child receives bear no consistent relation to his behavior . . . an apathetic or an anxious or a hostile individual is likely to result (**110**, p. 90).

Such statements must be regarded as hypotheses not facts; and, so far as the writer has been able to gather from a rather thorough search of the literature, except for two studies to be noted forthwith, they have never been subjected to any empirical test. A great deal of reading-in of adult emotion into the infant situation is evident in these interpretations. It is well known that only a few years ago the best "scientific" opinion favored scheduled feeding to promote healthy personality development in the child, and this view is still fairly widespread in the medical profession. Certainly, even very young infants learn to adjust their sleeping and eating habits to a schedule of regular feeding, and it would seem that this regularity might just as much serve to promote a feeling of security and confidence on the part of the child as any other feeling. Here, too, we lean to the belief that a specific discipline does not exert a specific, invariant psychological influence upon the child, and that its effect can be gauged only from study of the parental attitudes associated with its administration, of the child's constitutional endowment, and of the entire social-historical situation in which the emergent personality is located.

The two studies to which we have referred have not been fully reported, but the published account is such a choice specimen not only of difficulties of the particular problem with which we are concerned but of social science methodology in general, that we reproduce it nearly verbatim:

> Miss Trainham . . . reported on a study begun at the Merill-Palmer School. Two groups of children were chosen, one group on self-regulating feeding regimes, and the other not self-regulated. The parents of both groups were quite similar. . . . One group consisted of ten subjects and the other twelve. . . the results suggest in a very striking way that the self-regulated children are in all categories of the tests more advanced in their development. . . .
>
> Dr. Escalona had conducted similar studies . . . and had arrived at opposite results. She, too, had only worked with small groups. One group consisted of babies on self-regulated feeding regimes, and the other group consisted of babies who were on the conventional rigid feeding schedules. The group of babies who were more frustrated according to the standards of this group [i.e., those on scheduled feeding] showed more rapid maturation by test scores. Until she had heard Miss Trainham's report, she had explained her findings by the theory that the frustrated babies are forced to recognize the environment earlier (**109**, p. 33).

Nothing is so easy to manufacture as a theory to "explain" a particular empirical finding. To explain *all* the relevant findings is not so easy, and we have no simple solution to offer in the present instance. Our means of measuring personality are not very adequate; our knowledge of its genesis is even less adequate.

But what *can* be asked of social scientists today is a minimum degree of historical and cultural sophistication which would enable them to recognize the position which the human groups that they study occupy in the *real* social world to which they—and the social scientists—belong (as contrasted to the position ascribed to them in the artificial world of the social scientist's "control" laboratory. Unfortunately, the social scientist cannot ignore the history of his subjects prior to the time that he has them in his laboratory, as the physical scientist can usually ignore the previous history of the materials with which he experiments). The scientist's observations would, then, at least have the validity of descriptive statements about a designated portion of the real world; ultimately, after a sufficient number of such statements had been accumulated, a fairly complete description of the real world might be obtained. To generalize about psychological or social laws of all humanity, however, on the basis of limited observations on small, unrepresentative groups, can only lead to such predicaments as the one described above.

Weaning. The writer has been unable to find any experimental data correlating different forms and times of weaning with aspects of personality, but a brief discussion of the question may nevertheless be in order. As an event which, in Western society, usually occurs within a relatively short space of time, weaning can be classified as a trauma. However, in keeping with his tendency to universalize Viennese cultural norms, Freud did not see the culturally relative nature of weaning, and conceived of it, instead, as an invariably painful situation for the child:

> It looks . . . as if the desire of the child for its first form of nourishment is altogether insatiable, and as if it never got over the pain of losing the mother's breast. I should not be at all surprised if an analysis of a member of a primitive race, who must have sucked the mother's breast when he could already run and talk, brought the same complaint to light (**46**, p. 166).

This view that the infant has an "insatiable" yearning for the breast can only be regarded as an adult reading-in of emotion into the young infant. It is difficult to see why the purely physiological sucking capacity of the neonate can not be satisfied as well upon an artificial nipple as upon his mother's breast, or why a child who has never sucked his mother's breast need ever be weaned from it. Examination of the weaning trauma suggests that it may be regarded primarily as a product of a widespread (but not universal) social situation in which the child has come to associate not only nourishment but also warmth and affection with the mother's breast, and in which separation from the breast creates a fundamentally different social situation to which he must adjust.[8]

When weaning disrupts a series of satisfactions which the child has been receiving with regularity from infancy, it would be surprising if he did not react emotionally to this deprivation. One might suppose that late weaning would be more traumatic than early weaning, since in the former case the child would be more firmly accustomed to the satisfactions of breast feeding and hence more loath to forego them. Against this supposition the psychoanalyst counters that the developing child ego is better fortified to withstand the loss of the mother's breast at a later age; and that early weaning creates a craving for the oral satisfaction which has been denied,[9] resulting in a pessimistic or sadistic character, whereas late weaning results in self-assurance and optimism (**36**, p. 489). But this is all ex cathedra doctrine to the agnostic,[10] in the absence of scientific documentation. A contrary hypothesis is that weaning, whether it comes early or late, gently or severely, can not be said either to weaken or strengthen the child's sense of security until the entire situation in which the act takes place and the changes it produces in the parent-child relationship have first been appraised. Under some circumstances, it is conceivable that severe early weaning may help to produce a strong, secure ego;[11] whereas, under other circumstances, the same weaning might contribute to the development of an insecure, anxiety-ridden personality.

---

[8] Cf. Bernfeld: "In those children who receive their nourishment from a bottle given them by so many different people that they do not learn to recognize the mother (nurse) and source of milk, bottle, as a coherent whole loved object, the process [of weaning] is essentially changed" (**11**, p. 298).

[9] "One gets the impression that at developmental levels that do not afford enough satisfaction, the organism refuses to go further, demanding the withheld satisfactions" (Fenichel, **36**, p. 65).

Thumb sucking. As thumb sucking by infants and children has frequently been used as a diagnostic of oral deprivation in psychiatric and anthropological investigations, it is desirable to appraise the validity of this criterion.

Levy (**87**, **89**) has experimentally confirmed the young animal's drive to exercise its sucking function in various ways when this had not been adequately exercised in the normal process of nursing, and he has demonstrated that thumb-sucking in the human infant is *often* a response to sucking deprivation (**86**, **87**, p. 203). Roberts (**130**) has supported the latter finding.[12] Surveying these data, Sears concluded that

> On the whole, evidence supports Freud's view . . . that thumbsucking is a reaction to deprivation. It occurs not uncommonly when the nipple is withdrawn from an infant's mouth and its chief antecedent condition, as a habitual activity, is inadequacy of opportunity to suck extensively in connection with the eating process (**136**, p. 4).

There is absolutely no question that thumb sucking *may be* "a reaction to [oral] deprivation." But is important to point out that (1) oral deprivation *need not* result in habitual thumb sucking[13] and (2) when thumb sucking does occur, it is *not always* attributable to oral deprivation. As the latter point has not received the attention it deserves, some documentation for it may be offered.

---

[10] Some of the more devout have also expressed the opinion that the oral character is not clearly defined. See Fenichel (**36**, p. 488).

[11] Cf. Bernfeld: "The early, sudden, frustration . . . would offer the possibility . . . for the unfolding of an energetic ego" (**11**, p. 299). The logic behind this argument is that weaning serves to separate the mother-image which is originally merged with the child's ego, thus encouraging an early independent development of the child's ego. Bernfeld elaborates this view into a general hypothesis that peoples who wean their children early have a more well-defined sense of reality than do peoples who wean their children late (p. 294).

[12] In this study of 15 thumb suckers and 15 non-thumb suckers 7–8 months old, she observed that "non thumb suckers took a longer time for feeding than was taken by thumb suckers" (**130**, p. 7). Cf. also Ericson (**29**, p. 499) and Davis and Havighurst (**18**, p. 707) who note that proportionately three times as many middle-class as lower-class children, in Chicago studies, were thumb suckers; this can be correlated with the lesser opportunities which the middle-class child was afforded for sucking while nursing.

[13] Levy notes that "the submissive child is more likely to yield a finger sucking pleasure than an aggressive one, regardless of the quantity of oral tension" (**87**, p. 206).

Thumb sucking has been observed at normal delivery (**76**, p. 84), and Bernfeld reports that it is present in *all* infants during the first three months.[14] Thumb sucking has been noted for a minority of children in primitive societies where unlimited nursing on self-demand is the rule.[15] An interesting case in our own society concerns twins raised on a self-demand schedule:

> Each infant was allowed to nurse at any time as long as desired, and nursing periods ranged from three minutes to fifty-five. Sherry . . . never sucked her thumb while Peter sucked his almost constantly at four months and in moments of stress at thirteen months (**141**, p. 176).

In a Chicago study, Davis and Havighurst found that

> Negro middle-class children are treated much more permissively than the white middle-class children with respect to feeding and weaning. . . . Yet the proportion of Negro middle-class children reported sucking their thumb is almost the same as the proportion of white middle-class children so reported (**18**, p. 707).

Gesell and Ilg (**52**, p. 124), and Bakwin (**6**) have voiced their conviction that oral deprivation is a factor "of minor importance" in the general etiology of thumb sucking. The latter writes, "We have frequently seen babies nursed at the breast and seemingly satisfied in every way, who sucked their thumbs; and we have seen these babies develop later as well-adjusted children" (**6**, p. 100).

An adequate statement of the relation between thumb sucking and oral deprivation, thus, appears, to be: (1) thumb sucking is often the result of previous sucking deprivation (the child desiring to gain a pleasure which he has been denied); (2) it may also result from *excessive* gratifications (the child desiring to perpetuate or reconstitute a pleasure which has been gratified); (3) it may be associated with neither deficient nor excessive, but with entirely normal sucking experience. That thumb sucking may also result from factors extraneous to oral gratification is not to be excluded. It is clear, in any case, that no single-factor explanation of its etiology will suffice, and that it cannot, therefore, be used as an *unqualified* index of oral deprivation.

---

[14] He writes: " . . . all doubt concerning the regularly-established fact that all children suck their fingers even without being hungry is absolutely excluded during the first three months . . . " (**11**, p. 73).

[15] Mead's unqualified statement that "no primitive child whom I have ever seen or heard of sucks its thumb or fingers," which appears in an authoritative handbook of child psychology (**104**, p. 676) and which was repeated at the May 29, 1948 meeting of the Society for Applied Anthropology in Philadelphia, can only be deplored as an inaccuracy. Dennis has noted that of 288 Hopi children 2–9 years old, 13 or 6% were definite cases of thumb sucking (**22**, p. 181); Ward Goodenough informs me that cases of thumb sucking are to be observed among children on the island of Truk, despite unlimited breast feeding throughout the first year; and similar data could no doubt be obtained for many other tribes if sufficiently accurate observations were available. The more moderate thesis that finger sucking is comparatively rare among many peoples with adequate breast feeding of children remains, of course, in line with the evidence. Thus Kluckhohn notes that among the Navaho "Thumb-sucking is very rare, and we have never observed 'obsessive' thumb-sucking or a child asleep with its thumb in its mouth. When excited or disturbed, a child will not infrequently put the index or some other finger in its mouth" (**82**, p. 54).

"*Mothering*"

In a series of widely quoted publications, Ribble has waxed rhapsodic about the importance of adequate "mothering" to the development of sound personality and organic health. By "mothering" she means the tactile contact, cuddling, and stimulation of the infant "by one individual" (**127**, p. 459). Such mothering, she contends, "primes the breathing mechanisms," improves "the tone of the gastro-intestinal tract," "fosters functional integration," relieves "muscular tension," strengthens the infant's "sense of equilibrium," and gives him "a sense of security" (**127, 128, 129**). The sound of the mother's voice is also held to contribute to the physical and emotional well-being of the infant: "singing to the young infant soft, simple, rhythmical songs brings a sense of relief and security to his sensitive nervous system and feeds his stimulus hunger" (**128**, p. 45).[16] Finally, the infant is said to have a physiological need for sucking which helps organize his respiration, blood circulation, digestion, the development of facial muscles and of the brain, coordination of the eyes, "the age at which speech appears," etc. (**125, 126, 128, 129**).

It is unfortunate that such an influential writer has not attempted to draw a line between her empirical findings and her personal opinions. There is so much panegyric and so little satisfactory evidence in her writing that it is difficult for an impartial critic to evaluate many of her statements objectively. One gets the general impression that every infant hovers on the border of death, and that the balance in the direction of life must be tipped by adequate "mothering" and sucking. After a while, one wonders how it is that babies who are not fondled and sung to ever manage to survive at all.

That the sucking mechanism of the neonate is developed at birth is, of course, well known; and, in the sense that most mechanisms are made to be used, it may perhaps be said that a "drive" to exercise its sucking function exists at birth independent of the infant's need for nutriment Ribble asserts that this drive must be encouraged and satisfied if the infant is to survive and prosper, stating that "Fifty per cent of the 600 babies in our . . . study were definitely not 'self-starters' in sucking" (**129**, p. 628). Other authorities, however, have stated that "Sucking is present in all infants after 24 hours" (**137**, p. 64); and there can be little doubt that spontaneous thumbsucking, which is all but universa among infants, can serve to satisfy whatever physiological "need" may exist for exercise of the sucking function, where this is not satisfied in the normal process of nursing.[17] (See the previous section on thumb sucking.)

---

[16] Cf. Bernfeld: "Compairé emphasized strongly the significance of the mother's voice for the child's development; indeed, he becomes almost pious
p. 88).

It seems to us, therefore, that Ribble takes too hysterical a view of the neonate's organic and psychic resources, which are considerably greater than she allows and which enable it to thrive under a much wider range of care than the modern American pediatrician advocates to the American middle-class mother.[18] The same point can be made in regard to the supposed need for "mothering" during the first few months of life. The human infant is obviously incapable of surviving without adult care and protection and, if one wishes, one may thus speak of his organic "need" for such care. But the minimum needs thus defined are probably much less than the pediatrician is apt to prescribe.

The pediatrician's prescriptions are, in general, designed to recreate the sheltered position which the fetus occupies in the womb.[19] We do not wish to imply that this treatment of the infant is *not* good for it. But as Bernfeld has observed, there are many cultures in which infants are regularly subjected to treatment which, in almost every particular, violates the "foetusphile" prescriptions of the pediatrician, and in which they still grow up into normal, healthy adults:

---

[17] We are, however, by no means convinced that the activity of sucking is physiologically necessary for infant health, but only that it is the normal complement to it i.e., we believe that infants restrained from thumb sucking and fed intravenously could be potentially as healthy as normally suckled infants, and would probably even lose their sucking urge, just as it has been found that "if chickens are fed from a dropper for a long period of time after birth, they lose the pecking urge" (Levy, **92**, p. 659). A study by Sears on cup feeding of infants seems to support this view: "In general, the results indicate that breast feeding produces an oral drive, or at least strengthens it, and cup-fed children do not have as much oral activity during the first ten days of life as breast fed children." (Personal communication. A report of this study is in press in *Pediatrics.*)

[18] There are, of course, pediatricians of a broader outlook who are aware of the fac that infants in other times and places have managed to grow to healthy maturity withou having enjoyed the care and advice of pediatricians. Gesell and Ilg are among these "Fortunately the factors of developmental safety are so great that most children weathe the everyday crudity of our methods of management" (**52**, p. 131).

[19] "The manifold prescriptions which science gives the mother for the care of th new-born can be summarized in one phrase: warmth and protection from stimuli. Thi means that our care of the infant attempts to reconstruct and strengthen the psycho physical condition in which we found the foetus" (Bernfeld, **11**, p. 3). " . . . natura intuition ran ahead of science, and with swaddling clothes, and the cradle, which wa to imitate the bobbing about in the amniotic fluid, created an artificial substitute for th uterus which had been abandoned" (Alexander, **4**, pp. 123–124).

> If we combine these details into a composite description, we gain a conception of child rearing which represents the exact opposite of the foetus situation. The child . . . is bathed or washed in cold water; it is placed on a flat pad without coverings, or is swaddled or left to lie motionless (**11**, pp. 6–7).

Indeed, Bernfeld feels obliged to coin the term "foetusphobic" to describe the type of infant care prevalent in so many societies. There are cultures where "no attention is paid to the filth which accumulates and . . . the babies are caked with dirt" (**37**, p. 83), where the infant's head or feet are deformed, or where its nose, ears, or mouth, are subject to piercing and its body to scarification or other painful mutilation.[20] Until contrary evidence is forthcoming, there is no reason to believe that infants reared in a "foetusphobic" manner are any more neurotic or insecure, or differ in any significant aspect of personality from the "foetusphile" infants *because of any single type of treatment to which they are subjected.*[21] That foetusphobic upbringing will result in the death of a high proportion of infants is quite another matter; it may also ensure that only strong and healthy infants survive to maturity, whereas foetusphile practices encourage the preservation of infants with weaker constitutions.[22]

---

[20] Ploss (**121**, Vol. 1, pp. 288–338) gives an extended account of such treatment to which many primitive infants are subjected. The typical Flathead adult, e.g., had a head which was so flattened that the top of the skull, viewed in profile, was only one and a half to two inches wide. Mason expresses the belief that such deformation did not cause much pain "as it is done in earliest infancy, while the bones are soft and easily depressed" (**103**, p. 173).

[21] Greenacre has declared that "heightening of the sado-masochistic elements of the character . . . would seem to be the inevitable result" of bodily deformation, mutilation, or restraint in childhood (**61**, pp. 216–217). Such statements are based upon psychoanalytic principles derived from observations in Western society where mutilation carries sado-masochistic implications; that they reveal a lack of insight into the nature of culturally conditioned behavior is indicated by her reference to "the underlying sadistic aggression from which such frightful practices of restraint originate" (p. 216). The anthropologist must observe that the primitive parent who deforms the head of his child is not necessarily being any more sadistic than the Jewish parent who has his child circumcized or the American mother who lets her daughter's hair grow long and combs it daily, to the child's extreme discomforture: each parent merely conforms to culturally defined norms of religious behavior or of beauty. And if the anthropologist is led to protest against psychological explanations of the origin of cultural phenomena (see White, **147**), surely he may also protest against psychoanalytic interpretations of personality which are based upon cultural facts whose meaning has been misconstrued.

Our argument has been directed against an exaggerated insistence upon the need for tactile stimulation of the infant by the mother for the maintenance of its mental and bodily health. We do not mean to suggest that *some* contact and, more important, affection from the mother or her surrogate is not necessary if the infant is to grow into a "healthy" individual (i.e., one conforming to an ideal-typical norm held by Western psychiatrists). Spitz has shown that when nursery children six to eight months old have their mothers removed, some of the children develop depressive symptoms, which he attributes to their loss of love object (**138**). Ribble describes negativistic or depressive states arising in children deprived of their mothers which, apparently, have a similar etiology.[23] The high death rate in foundling homes, despite satisfactory hygienic conditions, has been cited as evidence of the child's need for attention and love if it is to survive; however, it is unclear whether love or individualized care is the crucial desideratum here (any uniform institutional regime which is not adjusted to the different constitutions of its members is apt to lose those members whose constitutions can not adjust to *it*).

A crucial experiment conducted by Dennis (**19, 20**) throws some light on the question of how much and what kind of attention the infant requires if its physical and emotional development is to proceed unimpaired. He raised twin girls from birth in conditions of extreme isolation. The infants were separated from each other by a screen and received only the purely minimum attention required for feeding and bodily care, being kept supine on their backs and not spoken to or fondled or played with or exercised in any way until they were seven months old. Because of their restricted practice, visually directed reaching and grasping, sitting alone and standing with help were retarded, but otherwise, the children "yielded during most of the first year a record of development not distinguishable from comparison records of infants in normal environments" (**20**, p. 157). The children responded affectionately to the experimenter during his brief visits, so that Dennis remarks: "Fondling is not necessary for the development of interest in, and every sign of affection for, the adult" (**20**, p. 156). His over-all conclusions agree with the ones we have drawn:

---

[22] Bernfeld has suggested that the foetusphile practices of modern cultures owe their origin "primarily to the economic-political interest of the nations in large populations" (**11**, pp. 7–8).

[23] Ribble (**127**, p. 461; **129**, p. 633). Cf. Levy (**88**).

> . . . normal behavioral development can occur in some infants when most of the first year is spent under conditions of minimum social stimulation and of very restricted practice. In other words, a large number of acts on the part of the adult which have been held by some people to be of importance may be dispensed with (**20**, p. 156).

Similar conclusions can be drawn from the investigation of two Viennese psychologists who compared the development of Albanian children raised under conditions of great physical restraint, with that of Viennese children (**17**). Until they are one year old, the Albanian children are bound securely to a wooden cradle customarily placed in the darkest corner of the room, often with a cloth thrown over their heads so that no light is visible. These children displayed poor muscular coordination, but once given an opportunity to practice, their performance improved rapidly so that it was clear no permanent retardation had been effected. Their social behavior, as measured by responses to the experimenters in a series of standardized tests, was equal or superior to the norms for Viennese children of the same age.[24] Identical observations could, undoubtedly, be made on the children of many primitive peoples who, tied securely to cradleboards during their first year of life, may experience comparatively little bodily contact or fondling by the mother.[25]

An important consideration to be kept in mind in evaluating the relation between the type of physical care which a child receives and its emotional development is the constancy with which the care is administered. It should be noted that in the cases of anaclitic depression reported by Spitz the children affected had each been accustomed since birth to the care and attention of their mothers. One wonders if the same reaction would have occurred had they (like Dennis' experimental twins) not been accustomed to such personal attention during their first half year of life.

Another point which would bear investigation is the extent to which care by a *single* adult is requisite for the development of "normal' personality.[26] Ribble, it will be recalled, makes this assertion but th anthropologist will suspect it may be predicated upon the monogamous nuclear family system of Western society. Need the child who grows u in a society with a strong, extended family maintain only one love objec during its formative years? Such a child cared for by siblings an aunts would hardly be expected to develop the abnormally withdraw personality common to so many orphaned and unloved children in ou own society. In short, Ribble and those who share her beliefs are no so much making an absolute judgment on the type of care which i necessary for sound personality formation as they are making a serie of recommendations which, implicitly, are based upon the nature of th child's social environment in the Western family and upon the natur of the family's position in Western society. Scientific investigations o desirable patterns of child rearing might proceed more successfully i the investigators were more conscious of this fact.

---

[24] Mead comments on this Albanian study that "we do not yet have any way of measuring the change in quality which is introduced when an activity is engaged in later than was organically possible" (**106**, p. 72), her inference being that some change in quality *has* been introduced. However, by operational rule, it is meaningless to talk of a change until it is possible to measure that change. Elsewhere, she suggests that children who have been bound to the cradle "will differ in manner of walking, occasions when they walk, significance of walking for the personality, type of balance disturbance to which the individual is subject, etc." (**105**, p. 673). These are more measurable indices, but they have not yet been demonstrated.

[25] A number of other peoples who subject their children to extreme restraint during the course of infancy are mentioned in the following section on restraint of motion.

Sphincter Training

The anal-erotic, anal-sadistic or compulsive character is the mos clearly-drawn picture in Freud's album of characterology. According t this picture, the infant derives a great deal of erogenous pleasure fron its bowel movements, and when this pleasure is inhibited by sever bowel training,

> The usual result is reaction formation . . . in the form of a drive to limit al his primitive, impulsive, pleasure-seeking activities. . . . In extreme cases . . one becomes parsimonious, stingy, meticulous, punctual, tied down with pett self-restraints. Everything that is free, uncontrolled, spontaneous is dangerou . . . (**114**, p. 747).

So widespread is the anal character in Western society that Roheim has argued that the entire culture "is based on sublimations or reactio formations of anal trends" (**132**, p. 283). Elaborating the Freudia theory, some psychiatrists have formulated a system which directl correlates adult character traits with specific forms of sphincter trainin in childhood:

> . . . it appears that the child who has been trained in continence of urine an faeces slowly and without pressure or punishment will yield control over hi eliminations without anxiety or resentment, learning to release without con

---

[26] Studies of personality development in orphanages and public institutions whic Freud and Burlingham (**42**) and Goldfarm (**55**), among others, have conducted, thro some light on this question, but the abnormal nature of this environment in our societ requires that generalizations from their findings be made with caution.

> flict; but a precocious or harshly coercive training that forces the child, before physiologically ready, to release to the outside demands, will set up resistance, accentuate retention as a defensive response and focus the child's behavior upon acquisitive or compensatory outlets for the denial of possession of his own eliminations (**39**, p. 22).

In line with this theory, Gorer (**58, 59**) has attributed many features of Japanese culture and character to the rigid bowel training of Japanese infants.

However, when one comes to look for empirical verification of the influence of sphincter training upon personality development, evidence which would compare the personality traits of two groups of children whose home environment was similar except for the time and stringency with which sphincter training was imposed, the data are too meagre either to substantiate or to disprove the Freudian thesis. It is unfortunate that so little attention has been paid to this particular problem, since it appears to be the area of infant care in which experimental disposition of psychoanalytic precepts could most readily be made.

Huschka (**71**) studied the reaction of 213 problem children at a New York hospital to the imposition of bowel training. She considered as "coercive" training which was started before 8 months or completed before 18 months, whereas bowel training initiated after 8 months and completed after 18 months was considered "adequate." She found that children who were "coercively" trained manifested "undesirable" responses such as constipation, fear, and rage more frequently than did children who were "adequately" trained; but there was no follow-up to correlate these responses with any permanent effects upon personality, and no control of normal children was used for comparison with the problem children. In a similar study, the same author secured data on the reaction of 215 problem children to the institution of bladder training (**72**).

Hamilton compared a group of 200 adults who recalled constipation or other anal-erotic traits in childhood with a group who did not in personality traits of stinginess, sadism, masochism, and concern for clothes; he found some evidence in favor of Freud's theory of the anal origin of these traits (**63**, pp. 467–472).

These fragmentary and inadequate studies on sphincter training and personality are all that we have been able to discover in the psychological literature. Ardent believers in psychoanalysis have shown a tendency to regard these and even less substantial data as confirmatory of their hypotheses, but a reading of the original studies does not incline an impartial critic to such optimism. Thus, one recent

"Despert found that early training produced an over-organized and compulsive personality" (**98**, p. 411). Reference to the original paper by Despert (**24**), however, reveals nothing to confirm this judgment, which can only be interpreted as a wishful reading-in of conclusions unwarranted by explicit data. The same writer, reviewing a series of studies by Fries, states that

> . . . the role of habit-training was studied thoroughly. That it influenced behavior was proved beyond doubt. . . . It can be assumed that early training is very likely to have ill effects on emotional development. . . . Untoward results are to be expected. What these results are in specific behavior has not yet been tabulated and set in conclusive terms. . . . Nevertheless, it is conceivable that eventually more specific cause-effect data may be forthcoming (**98**, p. 410).

Fries herself, however, subsequently drew a contrary inference from her work, stating that "any attempt to correlate age at which habit training was started with later personality traits, without consideration of the mother's personality type, would involve a large factor of error (**49**, p. 92).

In the light of the emotional argument in many psychiatric circles about the contribution of early sphincter training to the development of neurosis, a few words of caution as to the validity of this thesis may well be offered. Only fifteen years ago, an authority (**149**) writing in the standard handbook of child psychology advised that bowel training be started as early as six weeks, while another child psychologist stated that bladder control might profitably be initiated at 3 months (**135**, p. 74). The shift in opinion that has taken place in the past decade must be regarded as due to a middle-class fashion of the times rather than to any conclusive scientific validation of the preferability of later sphincter training. Against the swings of custom, of course, rational argument is futile.

A common reason which is advanced in support of the desirability of late bowel training is that myelinization of the pyramidal tract is not completed until the child is at least a year old, and therefore earlier habit training makes demands upon the child which its organism is not yet mature enough to fulfill voluntarily; i.e., the child "has to rely on the local conditioning of the rectal sphincter by reflex action rather than on voluntary control" (**71**, p. 305). The relevant neurological facts are as yet, however, inadequately known:

> . . . much is still to be learned about even the neurological aspect of the optimum age at which to begin bladder training . . . it is not yet possible to make any definite statement about the time at which myelinization becomes complete. . . . [F. R. Ford] concludes that the sequence in which the function of various structures develops undoubtedly corresponds to the sequence in which

> they become myelinated, but . . . some degree of function may appear in the absence of myelin. Also . . . Langworthy has suggested that myelinization may be the result of development of function rather than the cause (**72**, p. 256).

In opposition to the orthodox Freudian's emphasis upon strict bowel training as the causal agent of anal character formation, Horney has advanced a view we are inclined to accept which explains anal character traits as arising from a total social situation:

> . . . instead of relating these traits to the . . . "anal" sphere, one would understand them as a response to the sum total of experiences in the early environment. Thus the difference in point of view may be expressed . . . a person does not have tight lips because of the tenseness of his sphincter, but both are tight because his character trends tend toward one goal—to hold on to what he has and never to give away anything (**69**, p. 61).

Kardiner has pointed out that in a society which demands early bowel training of its childeren, "cleanliness and orderliness are forms of acquiescence to cultural demands, and are therefore forms of obedience" (**79**, pp. 43–44). Obedience, punctuality, cleanliness, hoarding, and the other "anal" character traits are certainly inculcated into the child by middle-class parents of capitalist society, this syndrome of traits being in part historically associated with the Protestant virtues which were so well rewarded during the rise of capitalism, and having also an independent culture-history of their own.[27] In these terms, the origin of the anal character can be understood as an adjustment of the individual on the level of personality to concrete economic and cultural conditions. The contrary contention advanced by Roheim, that these conditions are due to severe anal training, is historically untenable: a person can hardly hoard money, e.g., or be compulsively punctual if there is no money to hoard and no clocks by which to measure punctuality; and need it be added that money and clocks serve rational purposes and that their development can be explained historically without reference to the infant libido?

Outstanding among psychoanalysts in his recognition of the importance of social conditions in the channelling of libidinous forces, Fenich (35) has argued persuasively against Roheim's position on just th point. But even Fenichel holds that "anal eroticism produces the desi to collect something," though adding that "What is collected is dete mined by reality" (p. 89). To speak of a general instinctual "desire t collect something" as existing divorced from the material demands of specific cultural situation is, however, to invoke gratuitously a facto for which independent evidence can in no case be given. It is true, to b sure, that certain animals (e.g., rodents) hoard, but as the human anima does not exist separate from his culture, it is fruitless (untestable) t speculate about what might be his drives in a culture-less circumstance The more economical theory would appear to be to explain "anal' personality traits in terms of the cultural universe which gives rise t them.

---

[27] See Weber (**144**). Kroeber (**83**, pp. 600–602) furnishes a brief culture-historical sketch of the distribution of traits of cleanliness and order in time and space. He points out that the early Protestants, rebelling against the conservative and clean clergy of the 1400's, "washed little and looked on bathing as rarely needed and as likely to be dangerous to health. . . . This phase passed away with the slackening of religious intransigeance, with the growth of enlightenment, with the increase and spread of wealth and therewith of comforts, with the rise of the bourgeoisie. . . . It was the countries in which the new wealthy middle class became specially influential that took the lead in the new direction; hence the scrubbing of Dutch doorsteps and the proverbial Englishman with his portable bath" (p. 601).

Restraint of Motion

The problem of the child's emotional reaction to physical restrain has received a good deal of attention in the anthropological and psychiatric literature. Dennis (**22**), in particular, has advanced what appears t us to be an excellent solution of the problem, and his paper can be consulted by those who wish a fuller documentation of our position. Greenacre (**61**) and Levy (**92**) have also contributed summaries of the literature.

In the course of a report published in 1917 which described variou experiments they had performed on newborn infants, Watson an Morgan declared: "If the face or head is held crying results, quickl followed by screaming. . . . Almost any child from birth can be throw into a rage if its arms are held tightly to its sides" (**143**, pp. 166–167) This statement was repeated by Watson in many publications whic appeared during the following eleven years.[28] It gained credence fro an interpretation of the Freudian frustration-aggression thesis advance by some scientists as a universal principle of human psychology: tha aggression must always result from frustration.[29] Neither Watson' original finding nor the Freudian thesis, however, has (we believe) bee validated by empirical investigations.

A number of psychologists tested Watson's observation in experiments on newborn infants. Taylor found that in response to restraint o the arms no reaction occurred frequently enough to be called typical

---

[28] A list of these publications, appearing from 1917–28, is furnished by Dennis (**22** pp. 202–203).

[29] ". . . the existence of frustration always leads to some form of aggression" (**2** p. 1).

12 8.

and only 32 per cent of the infants tested made *any* response described by Watson as a reaction to restraint (**140**, pp. 79–80). Pratt, Nelson, and Sun found that in 58 per cent of the cases tested, infants remained passive in response to restraint of their arms: "no definite 'defense' or 'rage' reactions were found in any considerable number of cases" (**123**, p. 182). Stirnimann studied the reactions of 150 newborn babies to the tieing of mittens on their hands. Twenty-three babies got rid of the mittens; but most "showed no reaction unless the gloves were applied just before meal time."[30]

Prolonged restraint of infants is common and often very severe in cultures where it is the custom to swaddle the newborn child or tie it to a cradleboard during the first year of life. Among the Comanche, for instance, it is stated that "Children of six months when taken off the cradleboards had arms as stiff as though they had been pinioned" (**95**, p. 73). The Chinook child's head is bound to a cradle immediately after birth and the binding is not removed for a period of anywhere from three months to a year (**103**, p. 176). Throughout the first year, Albanian children are released from their cradles only once a day for a brief period of cleaning (**17**). The custom of infant restraint is very widespread:

> In both Americas the majority of aboriginal children were confined in some sort of cradle from their birth until they were able to walk about (**103**, p. 161).[31]
>
> . . . swaddling clothes were employed by the Jews, the Greeks and the Romans . . . and were used throughout the medieval period in all European countries. Swaddling did not begin to disappear in England until 1750 (**22**, p. 209).

Reports vary on the initial reaction of infants to the imposition of restraint in the form of cradling or swaddling. Often the child cries, kicks, or otherwise protests;[32] at other times the swaddling process evokes no complaint.[33] But it seems to be the unanimous testimony of observers that once the child has been placed in restraint, it accepts this restraint without complaint during the first few months of life. (It is difficult to imagine what else the child *could* do but submit, when it is securely bound and crying is not rewarded by any loosening of its bonds.) In fact, when removed from their cradles during the first six months, Navaho and Hopi children often cry to be put back in them.[34]

---

[30] We have been unable to consult the original reference (*Revue Fr. du Pediatrie*, 1937, **13**, 496–502) and so this account is taken from Levy (**92**, p. 660).

[31] According to Pettitt, the cradleboard was used in every cultural area in North America north of Mexico, except for the Eskimo and a few tribes of the Southeastern United States; and, with few exceptions, by every tribe within these areas (**119**, p. 12).

[32] Linton reports that Comanche children "always kicked and cried when they were being wrapped" (**94**, p. 472); the Albanian child at least occasionally cried when strapped into the cradle (**17**, p. 229).

[33] "Winifred de Kok . . . describes the swaddling of Italian children as an entirely happy proceeding . . . " (**61**, p. 212).

"The unanimous testimony of Hopi mothers affirms that when the infant is tied to the cradleboard on the first day of life the process does not cause the baby to cry" (**21**, p. 95).

As the child grows older and as it experiences increasing periods of freedom from restraint in the course of cleansing, play or exercise, it often will protest vociferously against the application of restraint.[35] On the basis of these data, Dennis has come to the following conclusions on the child's reactions to restraint, with which we are in agreement:

> [The young infant] . . . has no unique pattern of reaction to restraint. He reacts to all strong and persistent stimuli with restlessness and crying. . . . On the other hand, methods which achieve restraint of movement without steep gradients of pressure . . . do not produce restlessness and crying. . . . Whether or not restraint of movement will cause negative reactions in the older infant depends upon whether or not the restraint interferes with customary sequences which have been built up. *What is thwarting to the infant is not predetermined but depends upon post-natal events* (**22**, pp. 211–212, 216). [Our italics.]

In short, children who are accustomed to restraint do not experience it as a frustration; but once accustomed to freedom, the same restraint may become a frustrating experience. If they have never experienced restraint its imposition would, of course, generally be expected to provoke a hostile reaction.[36] These are substantially the same conculsions to which Levy was led;[37] and he also adds the important consideration that constitutional differences in the need of movement will influence the child's reaction to restraint.[38] These considerations must be known to psychiatrists who argue that restraint of motion is inevitably frustrating, for some of the same psychiatrists use restraint to effect the sedation of their over-excited mental patients. There appears, then, to be no uniform emotional reaction on the part of the child to a uniform type of restraint, but rather various reactions which differ primarily in accordance with individual constitution and previous experience.

---

[34] "Young [Navaho] babies will often cry to be put back into their cradles, and many of them do not sleep satisfactorily anywhere else" (**85**, p. 26).

"When the Hopi infant who has become accustomed to the cradleboard is kept off the board for a longer period than is usual he cries and does not go to sleep" (**21**, p. 96).

[35] "After about the sixth month . . . the [Navaho] infant apparently begins to feel the confinement a frustration and will wail to be released" (**85**, p. 26).

"Some Pueblo Indian infants at several months of age begin to object to being placed on the cradleboard. But it is important to note that the Indian infant of this age has been off the cradleboard each day for an interval sufficient to accustom him to manual play and to other activities" (**22**, p. 213).

[36] In this connection, Dennis reports that one Hopi mother who raised her daughter in freedom later tried to strap her to the cradleboard at five months, but the child struggled so strenuously that the mother finally relinquished the attempt, a "general experience of mothers who have given birth to a child away from the reservation, where cradleboards are not used, and who have returned to their native village when the infant was several months of age" (**21**, p. 98).

[37] " . . . after a child has experienced creeping or walking, a restriction of these activities will be felt as restraint. The modification of activities before they have been fulfilled may be easily achieved. . . . Once experience of free activity has taken place, the situation is quite different" (**92**, p. 659).

Infant "frustration" and "aggression"

It appears from the discussion of restraint (and, we can surmise, what is true of infant reactions to restraint may also to some extent be true of responses to other disciplines as well) that an experience which psychologists often designate as "frustrating" may in reality not be frustrating at all to the infant. Two outstanding factors enter into the explanation of those cases in which infant restraint was not experienced as frustrating: (1) the previous experience of the infant conditions it to accept restraint as normal, and (2) restraint is only part of a total situation in which the child finds itself and which contains many positive rewards (counterbalancing, apparently, whatever negative features may also be present). Thus, for example, the restrained Navaho child is protected "from the painful outcome of motor activity, such as picking up a hot coal or getting in the way of adults" and, at the same time, may actually participate in the social life of the family at an earlier age than does the uncradled American child.[39] Contrariwise, there is no society,

---

Cf. Hartmann, Kris and Loewenstein: "early swaddling . . . does not stimulate aggressive response, since it does not interrupt an activity but prevents one" (**65**, p. 25). These authors, however, also repeat the unjustified assertion that "By and large, the child tends to react to restraint by some manifestation of aggression" (*ibid.*).

[38] " . . . the need of movement varies according to . . . innate differences, maturation, health, type of restraint, and emotional factors" (**92**, p. 662).

Fries also lends support to the importance of constitutional factors in determining the child's reaction to restraint. "Children differ in their Congenital Activity Type. The difference in responses of the quiet, moderately active and active infant have been described in full by the author . . . . Infants of these three types approach similar situations in a different manner . . . a restriction of movement that is frustrating to the active child may not be so to the quiet child, whereas situations which are real obstacles to the quiet type are easily overcome by the active type" (**49**, p. 86).

[39] " . . . the cradle permits babies who could not otherwise sit up unaided to assume for long periods a position other than that of lying down, out of touch with what is going on around them" (**85**, p. 25). It might be added that the notion of an Indian cradle as a comfortless prison deserves to be corrected. Nearly all cradles used by North American Indians contained dangling toys and objects to interest the child, foot rests suitable for exercising the legs, etc. (**103**, p. 162).

even those in which unlimited breast feeding and the most lenient sphincter training are customary, which does not regularly inflict some pain and frustration upon its young. Even childhood experiences which the psychologist and anthropologist view as ideally "satisfying" and "unfrustrating" actually contain frustrating elements.[40] As Goldfrank has pointed out, "Pressures of some sort are exerted during infancy even in those societies which exhibit the greatest leniency in nursing or toilet training, and some degree of permissiveness is present even when infant disciplines are manifestly severe" (**56**, p. 517).

We can, then, safely assume that infant experience is never uniformly "frustrating" or "satisfying" but always contains a greater or lesser admixture of both elements. The big scientific problem is how to estimate by objective means the extent to which any single situation, as well as the whole of childhood experience, may be considered weighted in the direction of frustration or in the direction of satisfaction. Until this problem has been solved, the efforts of psychoanalysts and anthropologists to explain adult personality in terms of infantile frustrations and gratifications can not be regarded as scientific, since subjective judgment rather than objective procedure determines the weighting of infantile experience.[41]

Dollard and his associates attempt to define frustration objectively as "that condition which exists when a goal-response suffers interference" and, again, "an interference with the occurrence of an instigated goal-response at its proper time in the behavior sequence is called a

---

[40] E.g., the Hopi, Navaho and Okinawans have been cited as illustrative of societies which afford their children ideal, unfrustrating early experiences. The Hopi child, however, is whipped lightly the first time it is placed on its carrier's back, and it is a regular practice for the mother to place bits of food to the child's lips and then to withdraw this food (**56**, pp. 521–522); the child is also often kept in semi-darkness, a cloth being placed over the face-guard of the cradle (**23**, p. 109). Of the Navaho, Kluckhohn reports that "Teasing of children of nine months or a year and over is not infrequent—even by the mother. Sometimes it takes mildly sadistic forms. . . . There are also delays in the response to crying, even of pre-verbal children. A child of seven or eight months may cry fifteen or twenty minutes during the day before a busy mother picks him up" (**82**, p. 62). On Okinawa, we are told, "The enuretic child is cauterized by placing burning leaves against the skin. Most Okinawan adults carry scars from these fiery applications" (**107**, p. 394).

[41] John Whiting and Irvin Child have experimented, at Yale University, with the use of independent ratings by observers of the "custom strength" of various childhood disciplines, and of the "habit strength" to which they give rise. The effort to objectify thei procedures is commendable, but it runs the risk of substituting a cultural bias for personal one—i.e., all American raters or all American psychologists may agree i weighting a particular discipline as "severe" or "lenient" in terms of our values; Chines or Hopi observers, however, might weight the same disciplines differently.

frustration" (**25**, pp. 11, 7). According to this system, a discipline would not be frustrating if it did not interrupt any prior-established goal-directed activity, and by asserting that this is the case in regard to infant restraint, the evidence presented in the preceding section can be brought into line with the frustration-aggression theory. The theory accounts admirably for many experimental observations of infant behavior—e.g., that strength of crying following feeding frustration correlates with the strength of hunger instigation (p. 28). With "frustration" defined, however, the problem is now pushed back to the determination of the strength of instigation and of what constitutes goal-directed activity. Both "instigation" and "goal-response" are defined abstractly,[42] but in practice their determination remains ambiguous in certain crucial infantile situations with which we are concerned;[43] and circular (if no aggression appears, no goal- directed activity has been interfered with) rather than independent reasoning tends to resolve the difficulty for proponents of the theory (as it tends to resolve difficulties for the proponents of most social science theories).

When agreement can be reached that a given experience *is* frustrating to the child, we are still not justified in assuming that an aggressive response is inevitably called forth by the experience:[44]

> In Freud's theory frustration as such is supposed to arouse hostility. Actually, however, healthy persons—children as well as adults—are well able to endure a considerable amount of frustration without any reaction of hostility (**69**, p. 67).
>
> There are any number of frustrations that do not evoke aggressive response in the sense of discharging hostility against a social object or its surrogates. There are, for example, a number of experiments in which animals are frus-

---

[42] "An *instigator* is some antecedent condition of which the predicted response is the consequence . . . . strength [of instigation] is measured by the degree to which the instigated response competes successfully with simultaneously instigated incompatible responses" (**25**, pp. 3–5).

"An act which terminates a predicted sequence will be called a *goal-response*. The goal-response may be defined as that reaction which reduces the strength of instigation to a degree at which it no longer has as much of a tendency to produce the predicted behavior sequence" (p. 6).

[43] E.g., we are told that even "birth may be an event interfering with responses previously characteristic of the child in its intra-uterine state" (**25**, p. 58).

[44] Although Dollard and his colleagues have been associated with the view that frustration inevitably leads to some form of aggression, one passage in their work suggests that this formula need not always apply to infants: " . . . frustration can conceivably occur during the birth process itself and at any time thereafter. This need not be the case, however, with aggressive responses as here defined. The first reactions to frustration may indeed be of a random character and may lack that destructiveness which is here posited for aggression" (**25**, p. 11).

trated and in which such aggression does not occur. . . . The sucking frustrations in infancy cause finger sucking or sucking of other objects . . . rather than increased aggression. There is no proof that the so-called weaning traumas of infancy cause more aggression or even more phantasies of hostility . . . than in other children. The same may be said of all those frustrations that have to do with bowel and bladder control (**93**, p. 264).

The procedure of some recent analysts of "national character"[45] in dealing with this dilemma of what is frustrating to the child is particularly objectionable from any scientific point of view. It consists of making judgments as to the nature of adult personality and then looking for infant disciplines which can be regarded as having the requisite frustrating or gratifying effect for the production of this personality; or, vice versa, of making judgments about the nature of childhood disciplines and then looking for evidence of personality traits which might be fitted to them. A causal connection is then posited, but not proved. At no point is any search made for evidence which might disprove the thesis, nor are any objective rules laid down which will allow a second party to follow the same procedure independently and achieve the same results. (The analyst is thus safe in his lair, but the progress of science is scarcely advanced thereby.)

One result is that the same childhood experience is arbitrarily read as having one significance for personality formation in one society and the opposite significance in another[46]—the cradleboard in the Southwest e.g., is said to have induced "passivity," but among the Plains Indians, "aggressiveness." Or the same experience *in the same society* is interpreted as having two different significances for personality formation by different observers. Thus Erikson argues that the cradleboard of the Sioux "because it limited the earliest muscle expression may have left a certain inheritance of unchannelized energy to the teething period," and contributed to the "aggressiveness" and "sadism" of the adult Siouan (**30**, p. 139); whereas "Belden credits the cradleboard for the early development of stoicism in Siouan children, adding that the habit of leaving them bound in these cradles for hours on end, regardless of their crying, teaches them to bear life calmly" (**119**, p. 12).

HOW IMPORTANT IS INFANCY FOR THE STRUCTURING OF CHARACTER?

In previous sections, we have discussed some of the known facts about the relation between specific disciplines of child rearing and the development of personality. The present section surveys the over-all problem of the importance of the period of infancy in comparison with childhood for the shaping of personality. If it can be shown that major structuring of personality takes place after the first year or year and a half, clearly the contribution of infant experience in general and of these disciplines in particular to subsequent personality formation has been exaggerated by some theorists. We believe that this is the case, and will call attention to some of the evidence that had led us to this conclusion. The picture, however, is not entirely clear, and the honest investigator must confess that apparently discordant facts are faced at every stage of the inquiry. These facts must be reconciled, and not ignored, before any solution of the problem can receive general acceptance. In this predicament, we may draw consolation from the remark of one student in this field that "If a psychologist of personality had to limit his discourse to theories that were securely proved he would have nothing to recount. In his realm there are no certainties" (**116**, p. xii).

---

[45] Particularly Gorer (**58**, **59**), Erikson (**30**, **31**, **32**), and La Barre (**84**).

[46] On theoretical grounds, to be sure, this is perfectly possible; our criticism is directed against the failure to enunciate any rules which guide interpretation.

The biologic orientation of Freud's psychology and his concept of unconscious mental processes naturally led psychoanalysts back to the earliest period of infancy in their search for factors influencing personality development.[47] This movement was carried to its logical extension by Rank (**124**) who viewed the act of birth itself as the basic cause of future anxiety; since all infants experienced the "trauma" of passage through the vaginal canal, every child had some anxiety; especially prolonged or difficult births might produce more severe anxiety states. Greenacre (**60, 62**) has recently revived this thesis, with some modifications.

It is known that the unborn foetus responds physically to certain stimuli from outside the womb and to some behavior of the mother,[48] but, of course, this furnishes no evidence as to the existence of any correlative emotional or psychological responses in the intrauterine

[47] The approach of biologists to the study of personality may be illustrated by Stockard: "The personality of an animal . . . depends primarily upon its nature or constitution. . . . The biologist very well knows that the most important part of individual existence for man is passed and over before birth" (**138a**, pp. 21, 26). The concept of the unconscious leads in the same direction, as Bernfeld has observed: " . . . if one completely abandons . . . the equating of conscious and mental, then there is no other way than to trace the mental phenomena back to birth. Perhaps sometime later we may . . . penetrate beyond birth into the foetal being" (**11**, p. 212).

[48] " . . . following a hard day's work by the mother frequently the child in utero becomes so active, kicking and changing its position, that it is impossible for her to sleep. . . . The presence of fatigue toxin would seem, then, to produce a response in the baby through the mother's blood supply. . . . The response of irritation and restlessness suggests the experience of discomfort for the foetus" (**81**, p. 185).

state. (But see the apparently contrary view expressed by Dunbar, 28, p. 156.) Freud himself was critical of Rank's thesis, noting that "no trustworthy investigation has ever been carried out to determine whether difficult and protracted birth is correlated in indisputable fashion with the development of neurosis" (**47**, p. 96); he added, however, that the question was still open. It has been argued against Rank that "individuals who have been delivered by Caesarian section do not, as adults, differ temperamentally in any easily discernible way, from persons who have been normally born" (**110**, p. 87). However, this statement by Mowrer and Kluckhohn is, apparently, not based upon any empirical study.[49] The only such study of which we know was made by Pearson who compared the behavior of children whose birth took less than six hours with children whose birth took more than fourteen hours. His conclusion was:

> There seems to be a difference between the behavior of children whose birth-time was brief and those whose birth-time was long. The former presents somewhat more the picture of the "neurotic" child than the latter, which might indicate that a short birth was a more traumatic experience than a long one. But the duration of birth itself cannot be held responsible solely . . . . The child whose birth is short more frequently is the youngest and had very indulgent parents, while the child whose birth is long is the eldest, the oldest child of the sex or the only child (**117**, p. 288).

It can be seen that in this case, the relation between duration of birth and anxiety or neurosis was the reverse of that postulated by Rank. But the data available at present are clearly too meagre to allow an empirical resolution of the question, and opinions must be formed instead on the basis of one's general theoretical orientation and on inferences drawn from data that do not bear directly on the subject.

The same can be said about the view of recent proponents of a rooming-in plan for the hospital care of newborn babies that the common hospital practice of separating infant from mother during the two weeks of the lying-in period "definitely interrupts . . . the initial sense of security . . . [and] predisposes these sensitive infants to anxiety" (**129**, p. 631). At a recent scientific meeting, two advocates of the new plan spoke of the "tremendous" and "enormous" adverse effect which separation from the mother during this period would have on the emotional development of the child.[50] Such statements reflect only the subjective judgment of their authors, as is evident from the very form in which they are cast; there is as yet no evidence to support or disprove these assertions (and perhaps that is why the emotions are so frequently resorted to in propounding them).

---

[49] Comparative studies of the emotional and nervous reactions of infants delivered by Caesarian and normal birth are referred to by Kenworthy (**81**, p. 181) who does not, however, cite any supporting data. He observes that "the Caesarian sectioned child is prone to be less sensitized—he cries less, is markedly less irritated by the contacts of handling, etc.—than the first-born child delivered through the birth canal." May one then suppose that the Caesarian child is *not* less sensitized than the *second-born* child?

Analysis of the neonate state has been too much colored by a naive reading-in of adult emotions into the infant which is comparable to the anthropomorphizing views of insect behavior once current among an earlier generation of biologists. "When a child comes into this world, he comes in lonely and he is afraid," one psychiatrist declares (**9**, p. 582); another physician writes that the newborn is "a frightened, primitive little creature" (**2**, p. 593). Especially in psychoanalytic literature, it is often "stated quite confidently that the child at birth is capable of profound and intense feeling" (**75**, p. 752). We agree wholeheartedly with Hendrick's comment that these descriptions of the neonate represent "far more projection of analytic theory and adult passions than scientific observation" (**66**, p. 33).

To some extent, of course, inference is the only possible means of approximating to an understanding of the psychological condition of the newborn infant. But this inference should rest upon concrete observation of neonate behavior rather than upon preconceptions as to the nature of the neonate state or extrapolations from a later period. The best observations we have to date indicate that the newborn infant knows neither anxiety nor confidence, fear nor happiness, but exists in an affectless and presumably consciousless state. Its behavior is characterized principally by a great amount of undifferentiated mass activity (**74**, p. 39). Its "resistance to many types of stimuli is very high; even during handling the infant will often fall asleep" (**137**, pp. 74–75); most of its day is spent sleeping. "When an infant below four or five days of age is dropped one or two feet it frequently shows no perceptible response. . . . This is also true for so-called "pleasurable" stimuli, such as stroking or petting, to which many newborn infants show no overt reaction" (**137**, p. 145). Efforts to establish a conditioned response in the neonate have met with failure or the most equivocal results.[51]

---

[50] We refer to addresses given by Margaret Mead and Edith Jackson at the Spring Meeting of the Society for Applied Anthropology in New Haven, May 30, 1947. Cf. Jackson *et al.* (**74a**) where more temperate claims for the virtues of the rooming-in plan are made.

[51] Marquis (**99**) contends that she set up a conditioned response of foodtaking reactions to the sound of a buzzer in 7 out of 8 newborn infants during the first ten days of life. Pratt, however, has rejected her claim as invalid on technical grounds (**122**, p. 239) and, after a thorough review of the literature, concludes that "There is some evidence

Nevertheless, it has been observed that "research workers, as well as nurses and parents, are inclined to recognize personality, in terms of individual characteristics, as early as the first day of life" (**78**, p. 4). Jones and Burks are probably correct in attributing this early recognition of individuality to "differences in muscle tone, motor activity, sensory thresholds, [and what is interpreted as] emotional expressiveness . . . " (**78**, p. 5). Certainly there are important physical differences between infants at birth "in the maturity of the nervous system, in the way frustrations are overcome, in the response to the Moro test, and in the amount of activity" (**48**, p. 226). Fries divides newborn infants into three types, in accordance with the amount of activity displayed: "the quiet, the moderately active, and the very active" (**48**). Just how such physical differences correlate with later personality remains uncertain,[52] although offering a promising field for investigation. Whether one designates the physical characteristics of the neonate as its "personality" or whether one chooses to restrict the term to emotional and psychological phenomena which arise only in later months is purely a matter of definition.[53]

In the same ill-founded but influential article referred to previously, Watson and Morgan declared that the newborn infant evidenced three basic emotional reactions belonging to "the original and fundamental nature of man: fear, rage, and love" (**143**, p. 165). The "fear" response consisted of "a sudden catching of the breath, clutching randomly with the hands . . . blinking of the eyelids, puckering of the lips, then crying"; it was called forth by falling, loud sounds, or a sudden push (p. 166). The "rage" reaction, produced by restraint, has been discussed in a previous section of this paper (*supra*, pp. 21–24). The "love" response was evoked by "stroking, or manipulation of some erogenous

---

that certain responses may be conditioned to experimental stimuli during the neonatal period. The responses are difficult to establish, highly unstable, and cannot be set up in all infants" (p. 240). His view is echoed by Wenger. "Conditioning in the neonate is unstable and not easily obtained . . . anticipatory mouth opening occurs at least by the second month, but evidence of conditionability before that time is equivocal" (**145**, p. 85; cf. **146** and Munn, **112**, p. 376).

[52] Fries declares: "There seems to be a high correlation between amount of activity [in infancy] and ability to meet a new situation or to overcome an obstacle, deprivation or frustration" (**48**, p. 228). Support for her position comes from Davis and Havighurst's study of 200 Chicago children. They found that "The children who were most active when young, compared with those who were most quiet when young, tend to be: most active now, most punished, more aggressive, and less neat" (**18**, p. 706).

[53] Allport, e.g., holds that "the newborn infant *lacks* personality. Although many of its determinants are congenital, personality as such is not inherited. Only when the original stream of activity meets the environment, acting upon it and being acted upon by it, do the first . . . incipient traits emerge" (**5**, p. 122).

zone, tickling, shaking, gentle rocking, patting" (**143**, p. 167). From these three initial emotions, the entire gamut of adult emotional expression was presumed to develop by conditioning and extension. It can be seen how Watson's theory complements the extreme psychoanalytic view of the neonate as an organism whose personality can be permanently structured by very early experience and conditioning, and how closely the writings of Ribble parallel this early statement.

Just as his view of the infant's rage reaction to restraint has been rejected, however, so must the remainder of Watson's theory be dismissed as unsatisfactory. It overlooked "the possibility that some emotional reactions might not be elicitable at birth and yet appear later on as a result of maturation rather than as a product of conditioning" (**122**, p. 235). Behavior which Watson claimed to have observed in response to specific stimuli was not confirmed by subsequent tests.[54]

The most convincing attack upon Watson's theory was made by Sherman (**137**, pp. 114–143) whose observations also illuminate the subject of adult reading-in of emotion into the child. He showed motion pictures of infant behavior to a group of graduate psychology students and asked them to name the emotion represented, the stimuli which produced the behavior having been deleted. No agreement was exhibted in naming the emotions. Similar conclusions were reached when judges were brought into direct contact with the infants (the stimuli which Watson used to evoke "fear" and "anger" reactions were administered to the infant behind a screen, and the screen was then removed so that the students could judge the reaction). Sherman's conclusions deserve to be underscored: "*In estimating the emotional behavior of infants, people tend to read into it their own attitudes toward the expected reaction. . . . Frequently we ascribe to the child reactions of which he is not at all capable. . . . Most persons judge the emotional behavior of an individual in terms of the stimuli which have produced the reactions*" (pp. 122–123, 142).[55]

---

[54] See, e.g., Sherman's observation quoted above on p. 30; and the work of Dennis cited in the section on restraint. Repeating Watson's test for "fear," Irwin dropped infants two feet; only two of 82 trials results in crying, no response was noted in 12 per cent and various limb movements were observed in 88 per cent of the cases (**73**, p. 168).

[55] One limitation should be noted in regard to Sherman's study. The behavior ("emotions") which he asked judges to identify—"hunger," "anger," "pain," and "fear"—was all produced by disturbing stimuli. Had he attempted merely to distinguish general states such as "excited" and "quiescent," the judges' agreement would doubtless have been near-unanimous.

A study by Goodenough (**57**) found that university students rated almost six times better than chance in judging emotions represented by photographs of a ten month old child, but at this age a greater ability to communicate (and detect) emotion from facial expressions would be expected.

As a result of Sherman's work, many child psychologists have gone so far as to abandon the word "emotion" in referring to behavior of the young infant. "In its place they advocate a purely descriptive account of the extent of organismic involvement in a response under definite conditions of stimulation, age of organism, and so on" (**122**, p. 236). The consensus of opinion of competent students, summarized by Pratt after a comprehensive review of the literature in this field, is that "It seems certain that *such well coördinated reactions as those implied by the terms love, rage, and fear, do not exist at this early age,* [i.e. in the neonate]" (**122**, p. 211, our italics). The present view regards emotions as developing gradually in the infant by a process of differentiation from generalized to more specific reactions (p. 208). A study by Bridges (**12**) of the emotional behavior of 62 infants in a Montreal hospital, for example, distinguishes only two amorphous states at one month—"excited" and "quiescent." At three months "distress" and "delight" are differentiated, and it is only subsequently that "anger," "fear," and "affection" are manifested. One may reject this particular developmental sequence or terminology, but the fact of differentiation from initially generalized responses and the importance of maturation[56] in the development of emotions seems to be generally conceded.

The point we would emphasize is that, in the light of the findings of child psychologists, there does not appear to be any justification for speaking of emotions such as "fear," "anxiety," or "love" in the infant during the first months of life, as is so frequently done. Only confusion can attach to the designation of the physical states of early infancy by emotional terms which derive their meaning from a later period of life.

Although, as has been stated, conditioning is difficult to effect in the first weeks of life, some learning—especially in relation to the feeding situation—is definitely established for the first month of life. Thus a study of the feeding response of infants at a New Haven hospital showed that

> . . . the human infant does, within the first weeks of life, modify his behavior in accordance with the external demands imposed upon him by the culture into which he is born. Infants on a three-hour schedule "learned" to be "hungry" at the end of three hours, those on a four-hour schedule showed less definite, but

---

[56] "The varying effects of a given external event as children grow older can . . . be noted in occasional shifts in emotional response. At about 5 to 8 months, for example, a hungry but hitherto relatively calm infant may show signs of anger if there is a momentary delay on his feeding after he has caught sight of the bottle. . . . Similarly, as a child's abilities increase with age, an event that previously frightened him may provoke anger, and at a still later time the event may provoke amusement" (**75**, p. 758).

certainly suggestive, evidence of having learned to wait four hours for feeding (**100**, p. 277).

During subsequent months, a whole range of conditioned responses can be developed, which evidence greater stability as the infant ages; "retention of conditioned responses developed in infancy in some instances have been observed over periods ranging up to 7 months" (**112**, p. 376).

Two interesting experiments performed upon rats by Hunt (**70**) and Wolf (**148**) lend considerable support to the view that infant experience is an important determinant of adult behavior. Hunt subjected one group of rats to food deprivation for a period of 15 days starting at the age of 24 days, and a second group to a similar period of deprivation starting at the age of 32 days. Both groups were then given unlimited feeding for five months, after which they were again subjected to a five day period of starvation, together with control animals who had not been deprived during infancy. The experimental rats in the 24-day group responded to the second period of deprivation by hoarding more than two and a half times as many pellets of food as did the control rats who had not been starved in infancy; but the 32-day group did not show any excess hoarding. Hunt interprets these results as showing that early infantile frustration has a greater effect upon adult behavior than does later infantile frustration (**70**, p. 359), a conclusion which would extend to the psychological sphere an observation that has been established in the biological (especially the embryological) sphere (see Stockard, **138**a, pp. 45, 134 ff.). This, of course, is the firm logic behind the psychoanalytic position.

Wolf's excellent experiment (**148**) is of equal interest. He impaired the hearing of one group of rats by taping their ears and the sight of a second group by taping their eyes for several weeks during infancy. When the restraints were removed, the animals learned to see and hear normally as adults. But placed in a situation where they had to compete with control animals for food in response to a light stimulus, the animals blinded in infancy experienced difficulty in seeing and failed to get the food; when a sound stimulus was substituted, the animals who had been deaf during infancy similarly failed. Animals whose functions were impaired for a period *after* infancy (and then restored), were not handicapped in the exercise of these functions during the competitive situation (pp. 33–34).

Apart from the consideration of the applicability to a larger animal of findings made upon a smaller one, a question raised by these experiwhich reactions to a specific infantile frustration can be generalized and transferred to other situations. Murphy expresses the belief that

> Studies on the influence of infantile frustrations upon subsequent drive and behavior structure . . . suggest a certain transfer and generalization to a genuine "personality configuration" in the adult, whether animal or man . . . we may say that continuity is no longer merely the continuity of an item; it soon becomes the continuity of a system (**114**, pp. 723–724).

Linton has suggested that certain generalized responses are easy to establish in the early years of life and hard to establish later, in contrast to specific responses which may be acquired at any time during the life cycle (**95a**, p. 116).

If correct, the facts of (1) infantile conditioning showing up in adult behavior and (2) being generalized to a "personality configuration," are not necessarily contradictory to the position we have taken previously. The main question remains: does experience during the first year structure personality so firmly that it can not be restructured by later experience, or is the continuity of personality to be explained rather by the continuity of later experience, and of the constitution, with that of the first year?[57]

Probably no one answer to this question can be given for the entire range of possible infantile experience. If one burns a die upon the infant's body, its stamp will endure throughout life. Perhaps there are intense and prolonged forms of experience during the first year which can similarly cast the infant's personality into a mold which cannot be reshaped by any subsequent events.

In the normal range of infant experience, however, we believe that events subsequent to the first year or two of life have the power to "confirm or deny" the personality of the growing infant, to perpetuate or remake it, depending upon whether the situation of later childhood perpetuates or alters the situation in which the infant was reared. Since this is one of our most crucial theses, and bears heavily on the problem of the relative importance of cultural as against instinctual factors in personality formation, we will devote some space to its documentation.

---

[57] " . . . several studies . . . demonstrate *persistence* of personality patterns (occasionally even in the face of active efforts on the part of parents or others to break up the patterns). . . . For illustration we may mention Washburn's study showing constancy of smiling and laughing during the first year of life, and a study by Jersild and Marky which reports a correlation as high as .79 to represent the constancy of 'fights and quarrels' of pre-school children from one year to the next. . . . Persistence of the trends may be due merely to persistent factors in the environment . . . [or] early influence or experiences which set the whole future course of development" (**78**, pp. 93–94).

Cf. Benedict (**10**), Goldfrank (**56**), Beaglehole (**8**), and Linton (**95a**, pp. 141–145) for theoretical formulations regarding the effect on personality of cultural continuity or discontinuity in childhood training.

The picture of the infant obtained from many psychoanalytic accounts has much in common with the medieval notion of the infant as a homunculus or miniature adult. A scientific picture of the infant during the first year of life, however, does not support this notion, as Allport makes clear:

> [During the first year] . . . there are certain conditions that prevent the formation of dispositions as stable as those to be formed later. There is, for example, a low degree of retentivity for conscious experiences. . . . Then too, though the infant learns rapidly, he also forgets rapidly. Habits . . . are readily lost. . . . The propensity for an all-or-none type of emotional activity prevents his learning discriminative affective responses or the development of a hierarchy of likes and dislikes. His capacity for conceptualization is slight. There are, then, plenty of reasons why personality should be less stable, less predictable, and less consistent in the early months of life than at any later time. It is indeed never again as unorganized and unstructured as in infancy (**5**, p. 130).

The following evidence suggests the vital importance of post-infant experience to the formation of personality:

1. The personality of adult Pueblo Indians is marked by "maladjustment" and "apprehensiveness," despite an infancy characterized by unlimited indulgence (**56**).
2. "Navaho Indian infants receive a maximum of protection and gratification, but the anxiety level among adults is very high—presumably because of post-infancy traumas" (**110**, p. 96). "Navaho children who leave the hogans calm and well-poised return at the end of the first school year nervous and tense . . . the theorists . . . claim too much for the earliest years and do not pay enough attention to later events and to the total situation in which the mature person finds himself" (**85**, pp. 68, 111).
3. " . . . the Kaska personality norm is introvert and the Haitian extrovert, although in both societies babies are fed when they so desire, receive no pressure to control elimination, are comforted and handled when irritable or playful, and suffer no discipline for emotional willfulness. The indication is that other criteria than those associated with physiological urges are important in personality structuring, and moreover, that these may lie in the childhood rather than infant experiences" (**142**, p. 575).
4. St. Thomas children of the lower economic class are reared with great indulgence: "Breast feeding is ordinarily continued for ten or twelve months and not uncommonly for as long as two years. No regular feeding schedule is held to; the child is fed when it cries. . . . The training of sphincter control is not begun until relatively late, often not until the age of three or four years and very unsystematically then." Nevertheless, "there typically emerges an insecure, suspicious, egocentric adult" (**14**, pp. 43, 89).
5. "The pathogenic effect of later impressions is firmly established, and it can be assumed without danger that a severe traumatic influence during the war, for example, might set up a traumatic neurosis in a person whose birth had been quite normal. Indeed, under a sufficiently strong traumatic influence, the healthiest of individuals can become neurotic" (**4**, p. 114).

6. Some American ethnic groups "were giving children more of the kind of warmth and consistent mothering which psychiatry and pediatrics have recently come to consider desirable than were the completely American groups. . . . The unspoiled maternal attitude has much to counteract, however, when it confronts the instability of depression years . . . a satisfying infancy does not necessarily compensate for economic deprivation in the next ten years, and there is increasing evidence that later gratification may go far toward offsetting the effects of early frustration" (**115**, pp. 656–657).

7. In most societies, male and female infants are treated substantially alike for the first year or two of life, but differences in behavior are increasingly manifested in childhood and later years. Whatever behavioral and psychological differences exist between adults of different sex, then, must be attributed to biological factors and the differential cultural conditioning of post-infant years.[58]

Faced with evidence such as that cited above, some of which they have contributed themselves, Mowrer, Leighton, and Kluckhohn attempt to salvage a part of Freudian theory by reasoning that

> Infantile indulgence probably does constitute the firmest foundation upon which, if later circumstances are reasonably favorable, a secure and confident adult personality can be developed. But it affords only a possible basis; it does not, in and of itself, promise fulfillment . . . infant experiences, while placing certain constraints upon personality, give mainly potentialities. . . . Whether these potentialities become actualized or not . . . depends upon later social conditions . . . only an external condition of adult life brings out fully a predisposition, the basis for which has been laid in the experiences of infancy and early childhood (**85**, p. 11; **110**, pp. 95–96).

Assuredly, this formulation is a vast improvement over the simplicist theories of the relation between infant discipline and adult personality which are still in vogue. It is reminiscent of the compromise between the instinctual Freudian and the social-situational views of anal personality genesis which Fenichel advanced (**35**). As Fenichel spoke of the anal "desire to collect something," so these authors speak of adult "predispositions" and "potentialities" as "determined" by infant experience. Perhaps they are correct, but as there is still no way to measure a "predisposition" before the eventual empirical psychological *disposition* has been made, we wonder if there is not more mystification than meaning in the use of such terms. If infantile indulgence does not necessarily produce a confident personality, and if later gratification may offset earlier frustration, the range of "potential" personality is, in any case, too large to be fruitfully delimited on the basis of present knowledge. Instead of introducing teleological terms, it would seem preferable merely to describe the nature of organism and environment as precisely as possible at all stages in their history of interaction—infancy, childhood, adolescence, and maturity—allowing for the likelihood that (a) different infantile experiences acting upon different constitutions may produce the same personality type, (b) the same infantile experience acting upon different constitutions may produce different personality types, (c) the same childhood experience following different infantile experiences may produce similar personality types, (d) different childhood experiences following the same infantile experience may produce different personality types.

---

[58] "Emotional differences between the sexes have not been observed during early life although boy infants are believed by many to be more difficult to manage than girls. Girls talk somewhat earlier than boys but in other respects mental development proceeds at the same rate in both sexes. The greater aggressiveness of boys . . . becomes evident during the second or third year of life" (Bakwin, **7**, p. 46). Cf. Marshall (**101**, p. 12).

We conclude that the rigidity of character structuring during the first year or two of life has been exaggerated by many authorities, and that the events of childhood and later years are of great importance in reinforcing or changing the character structure tentatively formed during infancy.[59] Or one may substitute Horney's formulation: " . . . the sum total of childhood experiences brings about a certain character structure, or rather starts its development. With some persons this development essentially stops at the age of five. With some it stops in adolescence, with others around thirty, with a few it goes on until old age" (**69**, p. 152).

Discussion

It can be conceded that social scientists have failed to produce a definitive answer to the question of the relation between infant disciplines and character development, because of a general lack of historical and cultural sophistication, the difficulty of establishing the validity of the personality measurements employed, and the difficulty of isolating single factors for study. It is hard to see how the last obstacle, in particular, can be overcome. Social phenomena can not readily be subjected to the type of crucial experiment which enables the scientist to support or discredit an hypothesis. One has the feeling that social science theory, therefore, often moves in cycles like fads or persists like customs, for historical reasons, rather than progressing firmly in one direction because certain truths have been established by objective tests and can be built upon by the exercise of reason.

---

[59] Beaglehole (**8**) has proposed a useful distinction between the "primary" character structure formed in infancy, and the "secondary" character structure which may be formed in later years: "The integration-determining structure of needs and feelings constructed during infancy is generally reenforced and supported by the continued conditioning of the middle years of childhood and adolescence. . . . Hence this original character structure may very well be called the primary character structure of the person. Under normal circumstances and in relatively static societies, it will be the only character structure that the person possesses. In certain societies, however . . . the person may very well build or be forced to develop, either for permanent or temporary change, a secondary character structure which will enable him as a person to react dynamically and adaptively to a new set of major directives" (p. 149).

If the personality theory before us be of this nature, then the task we have undertaken in this paper of subjecting it to empirical test is indeed hopeless. Out of an obstinate spirit, however, let us proceed *as if* this were not the case.

Accepting, then, the data which have been reviewed as tentative markers on the psychological terrain, we are led to reject the thesis that specific nursing disciplines have a specific, invariant psychological impact upon the child. Instead, it appears that the effect of a particular discipline can be determined only from knowledge of the parental attitudes associated with it, the value which the culture places upon that discipline, the organic constitution of the infant, and the entire sociocultural situation in which the individual is located. In short, it is contended that personality is not the resultant of instinctual infantile libidinal drives mechanically channelled by parental disciplines, but rather that it is a dynamic product of the interaction of a unique organism undergoing maturation and a unique physical and social environment. This view is in substantial agreement with the position which Horney (**69**, p. 70) and Fromm (**51**, p. 286), among others, have advocated. Aside from meeting the facts of personality formation better than does orthodox Freudian theory, it has the advantage of being able to relate character structure to a concrete historical situation, which Freudian theory has been unable to do:

> As long as we assume, for instance, that the anal character, as it is typical of the European lower middle class, is caused by certain early experiences in connection with defecation, we have hardly any data that lead us to understand why a specific class should have an anal social character. However, if we understand it as one form of relatedness to others, rooted in the character structure and resulting from the experiences with the outside world, we have a key for understanding why the whole mode of life of the lower middle class . . . made for the development of this kind of character structure (**51**, p. 293).

In an illuminating passage, Murphy has pointed out that "there is no such thing as the 'true' or 'real' attitude, independent of the human situation. There are indeed many *potential* attitudes which are real, normal, human, and important, but none can appear and function except in a certain situation. Personality can be defined in terms of situations. . . . Personality could almost be defined as the integration of all the roles that a particular person has to enact" (**114**, pp. 876–877). If this is true of attitudes, must it not be equally true of infant experience, that no experience can be meaningfully isolated from the biological-psychological-cultural context in which it is embedded; that no discipline has an automatic, invariant emotional effect upon the infant? There is nothing new in this idea. Fromm (**51**, p. 293), Horney (**68**, p. 80), Kardiner (**80**, p. 224), Fries (**50**, p. 14), Lois B. Murphy (**115**, p. 654), Mowrer and Kluckhohn (**110**, p. 96) have all said the same thing at various times.

A particularly clear statement of this view was made by Pearson in 1931, after an intensive study of 72 problem children at the Philadelphia Child Guidance Clinic. The question which Pearson posed was whether the child's relationship to the parents, or the physiological events of its life were more important in shaping its personality. Under "relationships to parents," the attitudes of each parent toward the pregnancy, toward the child, and the position of the child in the family were determined; under "events," the health of the mother during pregnancy, the duration of pregnancy, length of birth, duration of nursing, duration of weaning, method of weaning, and method of bladder and bowel training were recorded.[60] His conclusion was that:

> . . . the parental attitudes exert a more important influence on the formation of the child's personality than the actual events. . . .
>
> There is no question that the physical situations of a child's life bear a definite relation to the formation of his personality, but the number of associated etiological factors renders it impossible to consider any single situation as causative. This study does indicate, however, the marked importance of parental attitudes. . . . The child forms most of his ideas about the world of humanity from his impressions of his parents, and their attitude to him . . . must affect the formation of his personality to a far greater degree than the length of his birth or the duration of his breast feeding. The events of a child's life enhance and crystallize the parental attitude toward him . . . and the child reacts to the parental attitudes through its association with the events of his life rather than to the events themselves (**117**, pp. 287, 290–291).

If this hypothesis is correct, the importance which Freud ascribed to traumatic events of childhood in structuring personality was greatly exaggerated. Several investigators have reached the same conclusion.[61]

---

[60] Cf. Davis and Havighurst (**18**) who, interviewing 200 well-sampled Chicago families, recorded similar data on child training and personality with the object of testing for any correlation between the two. Their conclusion was negative: "Very few pronounced relationships appeared" (p. 709).

[61] E.g., Sears: "Several sources of evidence indicate . . . . that Freud seriously overestimated . . . the importance of childhood sex aggressions . . . childhood [sexual] experiences are outlawed in our own culture and the child who has them, either willingly or unwillingly . . . is made to feel guilty or ashamed . . . the guilt and anxiety may in some instance lead to actual neurotic symptom formation. This would naturally be attributed to the experienced *event* rather than to the *emotions* involved" (**136**, pp. 36–37).

Assuming that the personality of the mother and her attitude toward the child are major factors in shaping the child's personality, the problem arises of when and how their influence is first manifested. Once the infant recognizes the mother as a distinct individual, let us say at the sixth month, significant interpersonal relations can certainly be said to have started. These earliest relations, however, should not simply be considered to affect the infant in the precise manner that an older child might be affected by the same relations. What may be called an "aberration affect" probably colors the experience for the infant until it is brought into sharper focus by maturation and learning.[62] Clearly, moreover, no fixed time can be set for the child's first conscious perception of its environment; and the annoying problem of unconscious perception remains. But to overemphasize the role of the unconscious in the first few months of life leads to the error which, we believe, Rank committed, and flies in the face of the descriptive data of infant psychology which have been summarized above. It is important to stress that the neonate and young infant is an immature animal organism, culturally "neuter" and psychologically uncommitted (to the extent that its constitution has not committed it), which can only slowly and with much parental effort and the gradual maturation of its faculties, be socialized. When Freud writes that little children "cannot help conceiving the sexual act as a kind of maltreating or overpowering, that is, it impresses them in a sadistic sense" (**45**, p. 55), he gives us insight into the Viennese but not into the infantile mind. Some anthropologists who regard the infant as a cultural being from the moment of birth—even of conception[63]—commit a similar error in failing to recognize that, although the neonate is instantly enveloped in a world of culture, it takes a considerable period of time until important elements of this culture are introjected, and the maturing infant remains, therefore, primarily a culture-less animal.[64] The ethnocentric, anthropo-

---

Cf. Linton: "... the potentialities of .... [traumatic] experiences for affecting the personality are probably determined quite as much by the attitudes of other individuals toward the incident as by any intrinsic qualities of the incident itself. Since these attitudes are primarily determined by culture, even this type of experience is culturally influenced" (**94**, p. 469).

[62] E.g. "... the very young child seems unable to recognize difference between facial expressions ... babies three months old may respond to scolding with definite signs of amusement" (**16**, p. 108).

[63] Statement of Margaret Mead at the May 29, 1948 session of the Society for Applied Anthropology in Philadelphia.

[64] How watery is the "culture of infants," can be judged from an interesting article which attempts to itemize the traits that infants of various ages are expected to acquire in the United States (Harlan, **64**).

morphic view is the most difficult error to guard against in studies of infant psychology.

Many writers speculate about a process of "empathy" or "emotional contagion" whereby the young infant senses the mother's attitude to it by means of subtle physical expressions, body tensions, or some less material signs.[65] At times the speculation hovers on the borders of mysticism; we have seen little evidence that such empathetic transfer takes place in the very first months of life.[66] The supposition is intriguing, however, and if it can be reduced to behavioral terms and confirmed, it would provide a vital missing link in the mother-child chain of personality communication. There is a good deal of evidence that subtle behavioral cues to maternal emotion are detected by the child in later months of life,[67] and that these cues may be more important in governing its character development than are the gross patterns of discipline which an observer may quickly note.

SUMMARY

This paper reviews some of the empirical data bearing on the theory that various features of infant care determine adult personality. Our conclusion has been largely negative, and we have been led to substitute a theory which emphasizes, instead, the importance of constitutional factors and of the total cultural situation in personality formation; the importance of post-infantile experience is also indicated. To the criticism that this paper stops where it should start, with the definition of specific situations which can be calculated to produce specific psychological effects upon individuals with certain constitutions, we can only agree.

---

[65] See Dunbar (**28**, p. 157), Escalona (**33**, p. 80), Greenacre (**61**, p. 209), Lundberg and Farnham (**97**, p. 305), Maslow and Szilagyi-Kessler (**102**, p. 84), Mullahy (**111**, p. 178), Plant (**120**, p. 96), and Sullivan (**139**, p. 8).

[66] Some of the most convincing evidence concerns the reported transfer of food aversions from nurse to infant. Escalona (**33**) notes that infants' acceptance or rejection of orange or tomato juice depended upon which juice the person feeding them liked or disliked, as well as upon the feeder's attitude to the infant. Similarly, Dunbar declares that "there is probably no point in prescribing for the infant any food substance for which the mother, nurse, or whoever is in charge of feeding, has a definite dislike" (**28**, p. 157).

[67] Dunbar, however, suggests that the psychological state of the mother can even be communicated to the unborn foetus: "The physiology of the mother is changed when she is under emotional strain and the effect of these changes is transmitted to the foetus through the placental circulation and in other ways" (**28**, p. 156). Cf. Greenacre (**60**).

BIBLIOGRAPHY

1. Abraham, Karl. *Selected papers.* London: Hogarth Press, 1927.
2. Aldrich, C. A. Highlights on the psychology of infancy. *Mental Hygiene*, 1946, **30**, 590–596.
3. Aldrich, C. A. The advisability of breast feeding. *J. Amer. med. Ass.*, Dec. 6, 1947, **135**, 915–916.
4. Alexander, F. *The psychoanalysis of the total personality.* New York: 1946.
5. Allport, G. *Personality.* New York: Henry Holt, 1937.
6. Bakwin, H. Thumb- and finger-sucking in children. *J. Pediatrics*, 1948, **32**, 99–101.
7. Bakwin, R. M., & Bakwin, H. *Psychological care during infancy and childhood.* New York: Appleton-Century, 1942.
8. Beaglehole, E. Character structure. *Psychiatry*, 1944, **7**, 145–162.
9. Bender, L. Mental hygiene and the child. *Amer. J. Orthopsychiat.*, 1939, **9**, 574–582.
10. Benedict, Ruth. Continuities and discontinuities in cultural conditioning *Psychiatry*, 1938, **1**, 161–167.
11. Bernfeld, S. *The psychology of the infant.* New York: Brentano, 1929.
12. Bridges, K. Emotional development in early infancy. *Child Develpm.*, 1932, **3**, 324–341.
13. Byington, G. M. Incidence of breast feeding in Detroit. *Psychosom. Med.*, 1945, **7**, 173.
14. Campbell, A. St. Thomas Negroes—A study of personality and culture. *Psychological Monographs*, 1943, **55**, No. 5.
15. Childers, A. T., & Hamil, B. M. Emotional problems in children as related to the duration of breast feeding in infancy. *Amer. J. Orthopsychiat.*, 1932, **2**, 134–142.
16. Collins, M. Modern trends in child psychology. In F. C. Bartlett *et al.* (Eds.), *The study of society.* New York: Macmillan, 1939.
17. Danzinger, L., & Frankl, L. Zum Problem der Funktionsreifung. *Zeitschrift für Kinderforschung*, 1934, **43**, 219–254.
18. Davis, A., & Havighurst, R. J. Social class and color differences in child-rearing. *Amer. sociol. Rev.*, 1946, **11**, 698–710.
19. Dennis, W. The effect of restricted practice upon the reaching, sitting, and standing of two infants. *J. genet. Psychol.*, 1935, **47**, 17–32.
20. Dennis, W. Infant development under conditions of restricted practice and of minimum social stimulation. *J. genet. Psychol.*, 1938, **53**, 149–158.
21. Dennis, W. *The Hopi child.* New York: Appleton-Century, 1940.
22. Dennis, W. Infant reaction to restraint. *Trans. N. Y. Acad. Sci.*, 1940, **2**, 202–218.
23. Dennis, W., & Dennis, M. G. Cradling practices of Pueblo Indians. *Amer. Anthrop.*, 1940, **42**, 107–115.
24. Despert, J. L. Urinary control and enuresis. *Psychosom. Med.*, 1944, **6**, 294–307.
25. Dollard, J., *et al.* *Frustration and aggression.* New Haven: Yale University Press, 1939.
26. Du Bois, C. Attitudes toward food and hunger in Alor. In L. Spier *et al.* (Eds.), *Language, culture, and personality.* Menasha, Wis.: George Banta, 1941.
27. Du Bois, C. *The people of Alor.* Minneapolis: Univ. of Minnesota Press, 1944.
28. Dunbar, Flanders. Effect of the mother's emotional attitude on the infant. *Psychosom. Med.*, 1944, **6**, 156–159.
29. Ericson, M. C. Social status and child-rearing practices. In T. M.

Newcomband E. L. Hartley (Eds.), *Readings in social psychology*. New York: Henry Holt, 1947.

30. Erikson, E. H. Observations on Sioux education. *J. Psychol.*, 1937, **7**, 101–156.
31. Erikson, E. H. Observations on the Yurok: childhood and world image. *Univ. Calif. Publ. in Amer. Archaeol. & Ethnol.*, 1943, **35**, 257–301.
32. Erikson, E. H. Childhood and tradition in two American Indian tribes. *Psychoanal. Study of the Child*, 1945, **1**, 319–350.
33. Escalona, S. K. Feeding disturbances in very young children. *Amer. J. Orthopsychiat.*, 1945, **15**, 76–80.
34. Faber, H. K., & Sutton, T. L. A statistical comparison of breast-fed and bottle-fed babies during the first year. *Amer. J. Dis. Child.*, 1930, **40**, 1163–1176.
35. Fenichel, O. The drive to amass wealth. *Psychoanal. Quart.*, 1938, **7**, 69–95.
36. Fenichel, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
37. Ford, C. S. *A comparative study of human reproduction*. New Haven: Yale Univ. Press, 1945.
38. Frank, L. K. The fundamental needs of the child. *Ment. Hyg., N. Y.*, 1938, **22**, 353–379.
39. Frank, L. K. Cultural coercion and individual distortion. *Psychiatry*, 1939, **2**, 11–27.
40. Frank, L. K. Freedom for the personality. *Psychiatry*, 1940, **3**, 341–349.
41. Freeman, M. Factors associated with length of breast feeding. *Smith Coll. Stud. in Social Work*, 1932, **2**, 274–283.
42. Freud, A., & Burlingham, D. *Infants without families*. New York: International Univ. Press. 1944.
43. Freud, S. Character and anal erotism. In *Collected Papers*, Vol. 2. London: Hogarth Press, 1924.
44. Freud, S. The infantile genital organization of the libido. In *Collected Papers*, Vol. 2. London: Hogarth Press, 1924.
45. Freud, S. *Three contributions to the theory of sex*. New York: Nervous and Ment. Dis. Pub., 1930.
46. Freud, S. *New introductory lectures on psychoanalysis*. New York: Norton, 1933.
47. Freud, S. *The problem of anxiety*. New York: Psychoanalytic Quarterly Press, 1936.
48. Fries, M. E. Mental hygiene in pregnancy, delivery, and the puerperium. *Ment. Hyg.*, 1941, **25**, 221–236.
49. Fries, M. E. The child's ego development and the training of adults in his development. *Psychoanal. Study of the Child*, 1946, **2**, 85–112.
50. Fries, M. E. Diagnosing the child's adjustment through age-level tests. *Psychoanal. Rev.*, 1947, **34**, 1–31.
51. Fromm, E. *Escape from freedom*. New York: Farrar, 1941.
52. Gesell, A., & Ilg, F. L. *Feeding behavior of infants*. Philadelphia: Lippincott, 1937.
53. Gesell, A. & Ilg, F. L. *Infant and child in the culture of today*. New York: Harper, 1943.
54. Glover, E. The significance of the mouth in psychoanalysis. *Brit. J. Med. Psychol.*, 1924, **4**, 134–155.
55. Goldfarm, W. Effects of early institutional care on adolescent personality: Rorschach data. *Amer. J. Orthopsychiat.*, 1944, **14**, 441–447.
56. Goldfrank, E. Socialization, personality, and the structure of Pueblo society. *Amer. Anthropol.*, 1945, **47**, 516–539.
57. Goodenough, F. The expression of

the emotions in infancy. *Child Develpm.*, 1931, **2**, 96–101.

58. Gorer, G. *Japanese character structure and propaganda.* New Haven: (Mimeographed), 1942.

59. Gorer, G. Themes in Japanese culture. *Trans. N. Y. Acad. Sci.*, 1943, **2**, 106–124.

60. Greenacre, P. The predisposition to anxiety. *Psychoanal. Quart.*, 1941, **10**, 66–94, 610–638.

61. Greenacre, P. Infants' reactions to restraint. *Amer. J. Orthopsychiat.*, 1944, **14**, 204–218.

62. Greenacre, P. The biological economy of birth. *Psychoanal. Study of the Child*, 1945, **1**, 31–51.

63. Hamilton, G. V. *A research in marriage.* New York: Boni, 1929.

64. Harlan, H. The culture of infants. *Social Forces*, 1944, **22**, 311–314.

65. Hartmann, H., *et al.* Comments on the formation of psychic structure. *Psychoanal. Study of the Child*, 1946, **2**, 11–38.

66. Hendrick, I. Instinct and the ego during infancy. *Psychoanal. Quart.* 1942, **11**, 33–58.

67. Hoefer, C., & Hardy, M. Later development of breast fed and artificially fed infants. *J. Amer. med. Ass.*, 1929, **92**, 615–619.

68. Horney, K. *The neurotic personality of our time.* New York: Norton, 1937.

69. Horney, K. *New ways in psychoanalysis.* New York: Norton, 1939.

70. Hunt, J. McV. The effects of infant-feeding frustration upon adult hoarding in the albino rat. *J. abnorm. soc. Psychol.*, 1941, **36**, 338-360.

71. Huschka, M. The child's response to coercive bowel training. *Psychosom. Med.*, 1942, **4**, 301–308.

72. Huschka, M. A study of training in voluntary control of urination in a group of problem children. *Psychosom. Med.*, 1943, **5**, 254–265.

73. Irwin, O. C. Infant responses to vertical movements. *Child Develpm.*, 1932, **3**, 167–169.

74. Irwin, O. C. The activities of newborn infants. In R. G. Barker *et al.* (Eds.), *Child behavior and development.* New York: McGraw-Hill, 1943.

74a. Jackson, E., *et al.* A hospital rooming-in unit. *Pediatrics*, 1948, **1**, 28–43.

75. Jersild, A. The beginnings of emotional behavior. In L. Carmichael (Ed., *Manual of child psychology.* New York: Wiley, 1946.

76. Johnson, Buford J. *Child psychology.* Baltimore: C. C Thomas, 1932.

77. Jones, E. Anal-erotic character traits. In *Papers on psychoanalysis.* London: Bailliere, 1923.

78. Jones, M. C., & Burks, B. S. *Personality development in childhood.* Washington: National Research Council, 1936.

79. Kardiner, A. *The individual and his society.* New York: Columbia Univ. Press, 1939.

80. Kardiner, A. *The psychological frontiers of society.* New York: Columbia Univ. Press, 1945.

81. Kenworthy, M. The pre-natal and early post-natal phenomena of consciousness. In E. Dummer (Ed.), *The unconscious.* Chicago, Knopf, 1927.

82. Kluckhohn, C. Some aspects of Navaho infancy and early childhood. *Psychoanal. & Soc. Sci.*, 1947, **1**, 37–86.

83. Kroeber, A. L. *Anthropology.* New York: Harcourt, Brace, 1948.

84. La Barre, W. Some observations on character structure in the Orient. *Psychiatry*, 1945, **8**, 319–342; 1946, **9**, 375–395.

85. Leighton, D., & Kluckhohn, C. *Children of the people.* Cambridge: Harvard Univ. Press, 1947.

86. Levy, D. M. Finger-sucking and accessory movements in early infancy. *Amer. J. Psychiat.*, 1928, **7**, 881–918.

87. Levy, D. M. Experiments on the sucking reflex and social behavior of dogs. *Amer. J. Orthopsychiat.*, 1934, **4**, 203–224.

88. Levy, D. M. Primary affect hunger. *Amer. J. Psychiat.*, 1937, **94**, 643–652.

89. Levy, D. M. Thumb or fingersucking from the psychiatric angle. *Child Develpm.*, 1937, **8**, 99–101.

90. Levy, D. M. Maternal overprotection. *Psychiatry*, 1939, **2**, 99–128.

91. Levy, D. M. *Maternal overprotection.* New York: Columbia Univ. Press, 1943.

92. Levy, D. M. On the problem of movement restraint. *Amer. J. Orthopsychiat.*, 1944, **14**, 644–671.

93. Levy, D. M. The hostile act. In T. M. Newcomb and E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1947.

94. Linton, R. *The study of man.* New York: Appleton-Century, 1936.

95. Linton, R. The Comanche. In A. Kardiner, *The psychological frontiers of society.* New York: Columbia Univ. Press, 1945.

95a. Linton, R. *The cultural background of personality.* New York: Appleton-Century, 1945.

96. Lorand, S. Character formation and psycho-analysis. In *Psychoanalysis today.* New York: Covici, 1933.

97. Lundberg, F., & Farnham, M. *Modern woman the lost sex.* New York: Harper, 1947.

98. Malcove, L. Margaret E. Fries' research in problems of infancy and childhood. *Psychoanal. Study of the Child*, 1945, **1**, 405–414.

99. Marquis, D. P. Can conditioned responses be established in the newborn infant? *J. genet. Psychol.*, 1931, **39**, 479–492.

100. Marquis, D. P. Learning in the neonate. *J. exper. Psychol.*, 1941, **29**, 263–282.

101. Marshall, T. H. Authority and the family. *Sociol. Rev.*, 1937, **29**, 1–19.

102. Maslow, A. H., & Szilagyi-Kessler, I. Security and breast-feeding. *J. abnorm. soc. Psychol.*, 1946, **41**, 83–85.

103. Mason, O. T. Cradles of the American aborigines. In *Smithsonian Institution Annual Report 1887*, Section 3 (Washington).

104. Mead, Margaret. The primitive child. In C. Murchison (Ed.) *Handbook of child psychology.* Worcester: Clark Univ. Press, 1931.

105. Mead, Margaret. Research in primitive children. In L. *Carmichael* (Ed.), *Manual of child psychology.* New York: John Wiley, 1946.

106. Mead, Margaret. On the implications for anthropology of the Gesell-Ilg approach to maturation. *Amer. Anthropol.*, 1947, **49**, 69–77.

107. Moloney, J. C. Psychiatric observations in Okinawa Shima. *Psychiatry*, 1945, **8**, 391–399.

108. Moloney, J. C. The Cornelian Corner. *Psychiat. Quart.*, 1946, **20**, 603–609.

109. Moloney, J. C. The Cornelian Corner and its rationale. In M. Senn (Ed.), *Problems of early infancy.* New York: 1947.

110. Mowrer, O. H., & Kluckhohn, C. Dynamic theory of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*, Vol. 1. New York: Ronald Press, 1944.

111. Mullahy, P. A theory of interpersonal relations and the evolution of personality. *Psychiatry*, 1945, **8**, 177–205.

112. Munn, N. L. Learning in children. In L. Carmichael (Ed.), *Manual of*

child psychology. New York: John Wiley, 1946.

113. Murphy, G., Murphy, L. B., & Newcomb, T. *Experimental social psychology*. New York: Harper, 1937.
114. Murphy, G. *Personality*. New York: Harper, 1947.
115. Murphy, L. B. Childhood experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*, Vol. 2. New York: Ronald Press, 1944.
116. Murray, H. A. *Explorations in personality*. New York: Oxford Univ. Press, 1938.
116a. Orlansky, H. Destiny in the nursery. *Commentary*, 1948, **5**, 563–569.
117. Pearson, G. Some early factors in the formation of personality. *Amer. J. Orthopsychiat.*, 1931, **1**, 284–291.
118. Peterson, C. H., & Spano, F. Breast feeding, maternal rejection and child personality. *Character and Pers.*, 1941, **10**, 62–66.
119. Pettitt, G. A. Primitive education in North America. *Univ. Calif. Publ. Amer. Archaeol. & Ethnol.*, 1946, **43**, No. 1.
120. Plant, J. S. *Personality and the culture pattern*. New York: Commonwealth Fund, 1937.
121. Ploss, H. *Das Kind*. Leipzig: T. Grieben, 1884.
122. Pratt, K. C. The neonate. In L. Carmichael (Ed.), *Manual of child psychology*. New York: John Wiley, 1946.
123. Pratt, K. C. *et al. The behavior of the newborn infant*. Ohio State Univ. Stud., Contributions in Psychol., No. 10, 1930.
124. Rank, O. *The trauma of birth*. New York: Harcourt, Brace, 1929.
125. Ribble, M. A. Clinical studies of instinctive reactions in newborn babies. *Amer. J. Psychiat.*, 1938, **95**, 145–160.
126. Ribble, M. A. The significance of infantile sucking for the psychic development of the individual. *J. nerv. ment. Dis.*, 1939, **90**, 455–463.
127. Ribble, M. A. Disorganizing factors of infant personality. *Amer. J. Psychiat.*, 1941, **98**, 459–463.
128. Ribble, M. A. *The rights of infants*. New York: Columbia Univ. Press, 1943.
129. Ribble, M. A. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*, Vol. 2. New York: Ronald Press, 1944.
130. Roberts, E. Thumb and finger sucking in relation to feeding in early infancy. *Amer. J. Dis. Child.*, 1944, **68**, 7–8.
131. Rogerson, B. C. F., & Rogerson, C. H. Feeding in infancy and subsequent psychological difficulties. *J. ment. Sci.*, 1939, **85**, 1163–1182.
132. Roheim, G. The study of character development and the ontogenetic theory of culture. In E. E. Evans-Pritchard *et al.* (Eds.), *Essays presented to C. G. Seligman*. London: 1934.
133. Roheim, G. The psycho-analytic interpretation of culture. *Int. J. Psychoanal.*, 1941, **22**, 147–169.
134. Roheim, G. *The origin and function of culture*. New York: Nervous & Mental Disease Pub., 1943.
135. Scoe, H. F. Bladder control in infancy and early childhood. *Univ. Ia. Stud. Child Wel.*, 1933, **5**, No. 4.
136. Sears, R. R. *Survey of objective studies of psychoanalytic concepts*. New York: Social Science Research Council, 1943.
137. Sherman, M., & Sherman, I. C. *The process of human behavior*. New York: Norton, 1929.

138. Spitz, R. A. Anaclitic depression. *Psychoanal. Study of the Child*, 1946, **2**, 313–342.

138a. Stockard, C. R. *The physical basis of personality*. New York: Norton, 1931.

139. Sullivan, H. S. Conceptions of modern psychiatry. *Psychiatry*, 1940, **3**, 1–117.

140. Taylor, J. H. Innate emotional responses in infants. *Ohio State Univ. Stud., Contr. in Psychol.*, 1934, No. 12, 69–81.

141. Trainham, G. Case history of the first thirteen months of life of a pair of twins breast fed on a self-demand schedule. *Psychosom. Med.*, 1945, **7**, 176.

142. Underwood, F. W., & Honigmann, I. A comparison of socialization and personality in two simple societies. *Amer. Anthropol.*, 1947, **49**, 557–577.

143. Watson, J. B., & Morgan, J. J. B. Emotional reactions and psychological experimentation. *Amer. J. Psychol.*, 1917, **28**, 163–174.

Received August 30, 1948.

144. Weber, M. *The Protestant ethic and the spirit of capitalism*. London: Scribners, 1930.

145. Wenger, M. A. An investigation of conditioned responses in human infants. *Univ. Ia. Stud. Child Wel.*, 1936, **12**, 7–90.

146. Wenger, M. A. Conditioned responses in human infants. In R. G. Barker *et al.* (Eds.), *Child behavior and development*. New York: McGraw-Hill, 1943.

147. White, L. Culturological vs. psychological interpretations of human behavior. *Amer. sociol. Rev.*, 1947, **12**, 686–698.

148. Wolf, A. The dynamics of the selective inhibition of specific functions in neurosis: a preliminary report. *Psychosom. Med.*, 1943, **5**, 27–38.

149. Wooley, H. T. Eating, sleeping, and elimination. In C. Murchison (Ed.), *Handbook of child psychology*. Worcester: Clark Univ. Press, 1931.

A NOTE ON GRANT'S DISCUSSION OF THE LATIN SQUARE PRINCIPLE IN THE DESIGN OF EXPERIMENTS

B. R. BUGELSKI
University of Buffalo

The suggestions made by Grant[1] that researchers might find a variety of uses for the Latin and Greek square arrangements of variables are especially to the point for psychologists concerned with learning problems in which a series of materials is learned in succession by the same subjects. Although Grant points out the similarity of principle between the Latin square and the traditional ABBA arrangements used commonly to control practice, fatigue, and other cumulative variables, there are some features of special interest in rote learning experimentation that were not brought out in Grant's general treatment.

The learning of a single list of nonsense syllables. The problem of which items in a single list are learned first, last, etc., is usually attacked by use of a list containing syllables of allegedly equal association value. Although individually the syllables may meet some criterion of low association value it may turn out that a given syllable is more or less easily learned depending on which syllable it follows. The writer has accumulated evidence to show that a given syllable has different learnability depending not only on its position in a list but also on what syllable it follows. Any claim made about positional effects based on the use of one set of syllables is open to criticism. The obvious corrective is either to use many different lists or to use the same syllables in different orders and arrangements so that each syllable follows every other syllable equally often, and in each and every ordinal position. Thus a six-syllable list could be expanded into a latin square consisting of six lists in which each column represents one arrangement showing the order in which the separate syllables appear, as follows:

1	2	3	4	5	6
2	4	6	1	3	5
3	6	2	5	1	4
4	1	5	2	6	3
5	3	1	6	4	2
6	5	4	3	2	1

In this square it is apparent that each syllable appears in each position only once and that it always precedes and follows a different syllable in each list (columns). In this way any combination of syllables that would tend to facilitate the learning of either or both is uniformly distributed among the subjects. Each syllable also appears in each position thus controlling any advantage that might accrue to a syllable from position alone.

---

[1] GRANT, D. A., The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, **45**, 427–442.

The learning of successive lists of syllables. In this type of problem a subject may be asked to learn one list of syllables or problems after another. The several lists should be thought of as comparable to individual nonsense syllables in a single list and the lists should then be arranged as above to avoid the possible cumulative effects of some one problem or list acting more positively in some other problem or list, thereby giving a spurious effect which might be attributed to position or to some other variable. In a proactive inhibition experiment, for example, it might make a difference whether list 4 follows list 5 or 6 or some other list, and if such other list is chosen at random, the results may be open to a cumulative effect error.

Odd-numbered vs even-numbered squares in experiments with nonsense syllables. Some of Grant's illustrative examples are taken from the field of agricultural experimentation wherein a number of variables are simultaneously counterbalanced by the Latin square technique. Under such circumstances it does not matter which item follows which in rows or columns since cumulative effects are not being considered. In such a circumstance, the Latin square shown by Grant

A	E	B	D	A
E	A	C	B	D
B	D	E	A	C
A	B	D	C	E
D	C	A	E	B

is apparently appropriate since each item appears in each row and in each column. It should be noted, however, that such an arrangement is not appropriate for nonsense syllables either in one list or in a series of lists because some of the items follow particular items more than once; thus if A has a specific effect on D, such an effect will occur twice because D follows A in column 1 and in column 5; again E follows C twice, B follows E twice, A follows B twice, etc. A little experimentation will demonstrate that it is impossible to arrange any odd-numbered set of items into a square without this difficulty, whereas there is no such problem with even-numbered lists. On page 439 in a foot-note discussing the Greco-Latin square, Grant states that a square of such a nature could not be developed with 6, 10, 14, and 18 unit sets of variables; this might be misconstrued to imply that experimenters should work with sets of lists or lists with odd numbers of items, whereas the contrary is true with the Latin square if cumulative positional effects are to be avoided.

Received December 8, 1948.

NOTE ON THE POSTMAN-STONE CONTROVERSY

JACK BERNARD
The Klein Institute for Aptitude Testing

Those readers of the *Psychological Bulletin* who are interested in the present status of Thorndike's law of effect have been provided with entertainment, if not information, by the year-long exchange of "notes" and "replies" between Postman and Stone (**4**, **5**, **6**). Since this exchange has reached the point where it consists largely of emotionally-charged words and phrases, it is perhaps time for a hitherto uninvolved party to risk an attempt to reconcile their divergent viewpoints.

Thorndike's law of effect, in its 1932 reformulation, served to split psychological opinion concerning the role of punishment in learning and to spur experimental studies. Postman's bibliography of 332 titles (**3**) attests to the vigor and enthusiasm with which that problem was attacked. However a half-century of extensive and intensive experimentation has served only to increase the divergence of psychological opinion relative to the effect of reward and punishment on learning. Both proponents and opponents of the law of effect have accumulated stores of research ammunition to justify their positions. Experiments may be cited to back the position that punishment helps, hinders or has no effect on the elimination of errors (e.g., **1**), and this is true of reward as well. There seems to be experimental evidence available as "proof" of any pet thesis which any individual may care to state as a "law," including the four logical possibilities suggested by Stone (**5**, p. 154).

Thorndike's law of effect in particular has not stood the test of time and research in terms of achieving the universality implicit in its formulation. Nor, in the opinion of the writer, can it withstand the test of common sense and everyday experience. If each of the trio involved in the present discussion were to burn his fingers in a new and unfamiliar type of fire, it is to be doubted if their later avoidance of that fire would have the indirectness of the creature who, according to Thorndike's "law," is "led by the annoying after-effect to do something else to the situation which makes him later less likely to follow the original connection" (**7**, p. 71). Human behavior, even the behavior of psychologists, tends to be somewhat more complex than that of the schematic sow-bug.

Does this mean that any effort to generalize within the important area of reward and punishment is hopeless? Not necessarily. It depends, to a considerable extent, upon how wide a generalization is attempted. It appears probable that any attempt at a "law" regarding "the effect" of reward and punishment on learning in general *is* foredoomed to failure. Available evidence suggests that reward and punishment do not have a single effect upon all types of learning under all conditions, but rather that at least four "effects" may be involved depending upon the specific situation. Tolman, Hall and Bretnall (**8**) have presented evidence for "laws" of emphasis, motivation and disruption, and evidence has also been produced which points to an informational value of such punishments as are applied in contiguity with errors (e.g., **2**).

It we recognize that any of the above effects may take preeminence in a specific learning situation, and that combinations of effects may operate simultaneously, we will have gone a long way toward reconciling divergent viewpoints and toward understanding the differences in results obtained experimentally. We will have recognized that the effect of reward or punishment in a learning situation depends on the specific nature of the reward or punishment, the specific learning situation involved, and, not least of all, the specific individual being rewarded or punished. Adherence to the law of parsimony, seeking for a single "truth" or "law" applicable to all situations, is admirable. But persistent, obstinate adherence to a generalization which has failed of its intended universality is illogical and unscientific.

Thorndike's generalization, popularized as the "law of effect," has a firm foundation in experimental research. However many other researches, just as competently planned, executed, and interpreted, yield results not in accordance with this generalization. To the writer, this suggests that a generalization as wide as the law of effect is not feasible. Perhaps several "laws of effect" are called for, each to describe results obtained within a narrower sphere. In the light of available data, neither battles of quotations nor battles involving the situation of selected confirmatory experiments serve a useful purpose. It may be more profitable to recognize the contradictory nature of the available evidence, recognize the possibility that at least to some extent

> "I am right, and you are right,
> And all is right as right can be,"

and seek generalizations which will *fit* all the available data more adequately than does the law of effect.

BIBLIOGRAPHY

1. Bernard, J. The specificity of the effect of shock on the acquisition and retention of motor and verbal habits. *J. exp. Psychol.*, 1942, **31**, 69–78.
2. Gilbert, R. W. The effect of non-informative shock upon maze learning and retention with human subjects. *J. exp. Psychol.*, 1936, **19**, 456–466.
3. Postman, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489–563.
4. Postman, L. Discussion of Stone's note on the law of effect. *Psychol. Bull.*, 1948, **45**, 344–345.
5. Stone, G. R. A note on Postman's review of the literature on the law of effect. *Psychol. Bull.*, 1948, **45**, 151–160.
6. Stone, G. R. Reply to Postman. *Psychol. Bull.*, 1948, **45**, 536–537.
7. Thorndike, E. L. *The psychology of wants, interests and attitudes.* New York: Appleton-Century, 1935.
8. Tolman, E. C., Hall, C. S., & Bretnall, E. P. A disproof of the law of effect and a substitution of the laws of emphasis, motivation and disruption. *J. exp. Psychol.*, 1932, **15**, 601–614.

BORING, E. G., LANGFELD, H. S., & WELD, H. P. (EDS.). *Foundations of psychology*. New York: Wiley, 1948. Pp. xv+632.

The Foundations of Psychology is the successor of the editors' previous elementary texts: *Psychology: a factual textbook*, 1935, and *Introduction to Psychology*, 1939 (both reviewed in THIS JOURNAL by the writer: 1935, **32**, 595–599; and 1939, **36**, 784–786). Although the editorial board remains the same, fifteen of the present eighteen contributors are new. It is no longer true, as it was for the previous editions, that most of the authors are "descendants" in the Titchenerian tradition. The present group of authors shows a great diversity of theoretical and professional background, and their writing well justifies the editors' statement that "In 1948 the important thing about the organism is not that it is conscious, but that it reacts to stimulation."

The sequence of chapters and their authors are as follows: the nature of psychology (E. G. Boring); the response mechanism (C. T. Morgan); response (C. T. Morgan); growth and development (L. Carmichael); feeling and emotion (W. A. Hunt); motivation (D. W. MacKinnon); learning (C. Hovland); retention and transfer of learning (C. Hovland); recollecting, imagining and thinking (T. A. Ryan); perception (E. B. Newman); sensation and psychological measurement (S. S. Stevens); color (F. L. Dimmick); visual space perception (F. L. Dimmick); hearing (E. B. Newman); taste and smell (C. Pfaffmann); somesthesis (F. A. Geldard); topographical orientation (D. G. Griffin); individual differences (A. Anastasi); heredity and environment (A. Anastasi); efficiency (M. E. Bitterman and T. A. Ryan); personality (L. F. Shaffer); personal adjustment (L. F. Shaffer); vocational selection (C. L. Shartle); attitudes and opinions (H. Cantril); and social relations of the individual (L. P. Crespi).

This is an impressive list of highly competent authors assigned to write on the broad fundamentals of general and applied psychology. The sequence of topics conforms to widespread opinion, and the various presentations are in general clearly and authoritatively written. The chapter on topographical orientation is the only one whose presence is apt to be seriously questioned. The subject is interesting, as are many others not included in the book, and well organized; but to the reviewer the topic has no valid claim for inclusion as a major item in an introductory text. Were it not for the fact that the chapter on "Response" is already probably overloaded with a miscellaneous series of topics, one would be inclined to say that topographical orientation belonged in part there and in part under "Hearing" and "Visual Space Perception."

It fell to Boring's lot to write the first chapter on "The Nature of

Psychology," the most difficult one to write, the least satisfying one in the earlier editions, and the one of most questionable value in any beginner's text. Except in this first chapter, there is almost no reference to consciousness in the volume, other than the casual use of the term in such expressions as "conscious pangs of hunger" and "the stream of consciousness." Boring, however, tells the reader emphatically that consciousness is a part of the subject matter of psychology, the other part being behavior. The individual authors nevertheless do not seem to justify such a statement by their treatment of the science. Certainly the student will be at a loss to find discussions of consciousness in the body of the text to compare or to coordinate with the accounts of behavior. Boring's statement that "only a hundred years ago you would have been told that psychology *is* the study of consciousness" (p. 5) will also seriously confuse the student unless he notes (p. 9) that "As late as 1910 . . . most psychologists . . . believed that physiology studied the body, that psychology studied consciousness and that introspection was the direct way of getting at consciousness." The place of consciousness in the psychology of the first chapter is further confused, for the reviewer and perhaps for professional readers, by the statement that "A science of psychology cannot, of course, be built up on any one man's experience, so most of consciousness for every psychologist is the experience of others." Presumably this experience of others is accessible to scientific study. In the discussion illustrating the nature of consciousness, the author includes the following statement: "One of the commonest illusions is seeing a single object as single with two eyes." If the object is single, seeing it as single should not be an illusion in the general acceptance of that term, although seeing it double might be. Ordinarily the term illusion is applied where there is a discrepancy between the "experience" and the "real object" or between the way in which the subject behaves in the experimental situation and in a standard situation where rulers, compasses, or tactual verifications can be employed. Perhaps, however, even for the author, consciousness is a scientifically negligible affair since he includes no definitions of such processes in the list of terms at the close of the chapter.

There are two additional points in this first chapter where adverse criticism seems to be in order: (1) Figure 2 purports to show the chronological and subject-matter relationships between the schools of psychology. It would take an undue amount of space to query this figure in detail. The only clear and unambiguous features, to the reviewer, are the dates. Surely then no beginner could interpret it and come out with anything approximating the historical relationships. (2) Under the heading of experiment as a part of scientific method, the author states that "Experiment is the observation of *concomitant variation* and the interpretation of the concomitances as causes and effects." Now the use of concomitant variations is an extremely important method of experimentation, but it is not the only one nor is the interpretation of "concomitances as causes and effects" as wise and safe a procedure as the statement implies. To interpret correlation coefficients or learning curves as cause and effect relations between the variables concerned can result in some strange conclusions. The experimental methods of agreement and difference, to use Mill's terms, deserve more specific recognition than they get under the author's term "control." Where one deals with the determination of the effect of benzedrine on intelligence scores or with the effect of the extirpation of one brain area upon behavior, it is confusing to try to fit the methodology to concomitant variation as though one must always deal with continuous functions.

Although the authors of the various chapters present highly competent and sometimes brilliant expositions of their topics, it is always possible for a reviewer to indicate ways in which the accounts suffer or ways in which they could be improved from his point of view. In the spirit of this observation, the following points are selected for mention: (1) In spite of the amount of space allotted to the nervous system and the receptors, no indication is given of the afferent nerves involved and but little mention is made of the effects of nerve lesions. (2) Neither in the chapter on psychological measurement nor in the one on individual differences can the student learn the simple statistical formulae necessary for his work or how to use them. (3) The chapter on color might have given less space to color equations and the color pyramid and more space to such psychophysiological topics as dark adaptation, integration, intensity-time relationships, etc. The section on "physiology of color vision" is notably weak. (4) Stevens, in the chapter on sensation and psychological measurement, writes as follows: "The sense organs start the messages along the nerves, the highways to the brain. When these messages merge at the higher centers of the nervous system, when they organize themselves and modify one another through interactions and associations, we call the result perception. But the bare messages themselves, isolated and apart from their mutual influences, we call sensations. We get sensations by analysis, by paying attention to certain aspects of our perceptions. . . . By selective attention he [the artist] analyzes the organized scene into patches of color and he translates these sensations into pigments on canvas." Now the reader has already (painfully) learned, in the chapter on the response mechanism, that the messages carried by the nerves are electrical changes. As Stevens states it, electrical changes are therefore the sensations, for example patches of color; and we get at these electrical changes by paying attention to aspects of perception which are the integrated electrical changes in the brain. The reviewer does not believe that Stevens means what he says; but he believes that the careful student will be confused by the account.

In the editorial preface to the volume, it is pointed out that the present book is about twice as large as the previous edition. "The strongest evidence of the rapid advance of psychology is the need of frequent revisions of textbooks. It is nine years since our last book, *Introduction to Psychology*, was published. In the meantime we have had a long war and a victory which psychological research helped attain. In this wartime research much new and valuable knowledge came into being. In addition there has been the more normal acquisition of facts as well as a clearly distinguishable change in point of view." Here we deal with matters of opinion. The reviewer finds no new point of view in the book and no chapters on topics which could not have been well discussed ten years ago. The reviewer does not believe that the advances made by psychology in the past nine years have been so extensive as to require a doubling of the size of elementary textbooks. Nor does he believe that wartime research produced more than a barely appreciable amount of new knowledge in psychology, although the applications of existing knowledge were varied and extensive. During the war, as before and after it, the fundamental handicap of applied psychology was the limited amount of basic knowledge available from general psychology. Such knowledge comes slowly. It is the reviewer's guess that the increased size of the present textbook over its predecessors is due far less to the growth of psychology than it is to the combination of changed contributors and the probable instructions given them on the lengths of their respective contributions.

In its current size, the justly famous BLW text steps out of the class of one-semester books and into the class of texts for full-year courses. Whether or not this is a pedagogical gain remains to be seen. However, to the student or the young instructor who masters the volume, there will come an excellent persepctive view of the science which will aid him to keep his thinking on a straight course through the confusing eddies and back waters of psychological activities in the post-war period.

WALTER S. HUNTER.

Brown University.

DENNIS, WAYNE (ED.). *Readings in the history of psychology.* New York: Appleton-Century-Crofts, 1948. Pp. xi+587.

Should the history of psychology be taught from original or secondary sources? If the student is to learn by eye, from the printed word, he will have to use secondary sources. The instructor could not provide fifty students with access to the original writings of, say, one hundred men. The books of the very great may be available in reprints and translations but not so often their crucial articles, except when the papers have been collected in anthologies. Dennis' anthology of sixty-one historically significant excerpts from forty-nine authors does us, therefore, a service. It enables the teacher to supplement one of the standard texts by samples of original sources. In this way many students will come better to comprehend the temper of some of the famous psychologists and the atmosphere in which they wrote.

Rand's *Classical Psychologists* of 1912 had a similar use, but it is out of date. History gets out of date. In 1912 the philosophers were much more important and the physiologists much less important than in 1948. Besides, 1948's past can now contain a little of what was only 1912's future.

Dennis does not, however, give enough material to make his book satisfactory as a sole text. It is conceivable that a thousand pages consisting of one hundred excerpts from eighty authors might do as a sole text, with the instructor's lectures filling in the interstices and setting up the dynamic interrelations, but I am not sure. Direct quotation is so uneconomical. What had to be labored in 1838 can be quickly understood in 1948. Johannes Müller's elaboration of his ten laws of specific nerve energies was necessary then, while thought still resisted his dicta, but not now, when what he was saying has become a commonplace.

The simplest way to indicate the content of this book is to give a chronological table of the forty-nine authors and their sixty-one excerpts. Here each mention of an author indicates a single excerpt, except when a parenthetical number indicates a joint authorship or more than one excerpt from the same author in the same period. In most cases readers will be able to guess the nature of the subject matter from the date and the author.

Antiquity: Aristotle (2)	2
1601–1650: Galileo, Descartes	2
1651–1700: Hobbes, Mariotte, Newton, Locke	4
1701–1750: Berkeley, Hartley	2
1751–1800: Mesmer, Thos. Young, Dalton	3
1801–1810: Thos. Young	1
1811–1820: Chas. Bell, Thos. Brown	2
1821–1830: Flourens, Jas. Mill	2
1831–1840: E. H. Weber, Joh. Müller	2
1841–1850: J. S. Mill, Braid, E. H. Weber, Helmholtz	4
1851–1860: Helmholtz, Fechner	2
1861–1870: Helmholtz, Galton	2
1871–1880: Wundt	1
1881–1890: Preyer, G. S. Hall, Galton (2), James, Ebbinghaus, H. H. Donaldson, Cattell (4)	11
1891–1900: Dewey, Titchener, Thorndike	3
1901–1910: Thorndike ($1\frac{1}{2}$), Woodworth ($\frac{1}{2}$), Binet, Simon, Pavlov, J. R. Angell	6
1911–1920: Watson, Hunter, Cannon, Terman, Köhler (2), Franz ($\frac{1}{2}$), Lashley ($1\frac{1}{2}$)	7
1921–1930: Yerkes, Carmichael, Barbara Burks, Lashley, Hull	5
	61

It is plain that Dennis has welcomed the philosophers with acute discretion, admitted the important physicists, turned definitely toward the physiologists, and accepted mental testing, while he excludes both abnormal and social psychology. Galton, Cattell as tester, Binet and Terman are here, but there are no signs of Charcot, Janet, Freud and the contemporary psychopathologists, nor of Le Bon, Tarde, McDougall and the modern American sociotropes. In this Dennis follows more than he leads, for this history which he documents is that part of psychology's history which is best known and has been most fully presented in texts.

The distribution against dates seems to me to be about right. Aristotle should come in, but the rest can be omitted until the seventeenth century. Rand gives fifteen philosophers before Hobbes to Dennis' one. It is jolly to have at hand that passage from Galileo on pitch and frequency. At the other extreme of time I pause to consider and assess my contemporaries. Dennis cites twelve living men, all except Köhler born in America. Dewey at eighty-nine is the oldest, Carmichael at fifty the youngest. The baby is Barbara Burks, who would have been forty-six this year. I doubt if she should have been included (the only woman in the list, too), but the others all seem to me to be justified examples of important current history. Discrimination in this matter is, of course, as invidious as it is necessary, but I am ready to support Dennis. Of the class of living APA ex-presidents over fifty years old, Dennis has included nine and omitted twenty-one. If we accept his exclusion of abnormal and social psychology from the project, then I think Dennis has done pretty well. I should add to his list, considering only these two: Tolman on reaction as meaning or on purposive behavior (as he takes the torch from E. B. Holt whom I should also want to work in), and Thurstone on mental abilities.

Now what we have here is sixty-one samples of historically significant contributions by forty-nine psychologists. If I were free to enlarge this book, I should keep all of these samples with five possibile exceptions. (1) I should leave out J. S. Mill on ethology and substitute Bain on laws of association, putting in Mill on the nature of matter, which is his theory of perception. I think it had more influence in the thought of that time. (2) Wundt is indeed a puzzle. What should one quote? These three pages on psychological methods from the fifth edition of the *Physiologische Psychologie* (1902) are quite inadequate. You get a fuller discussion of this sort in the introduction to his *Sinneswahrnehmung* (1862), where Wundt might be said to be 'founding' the new psychology, and you get something more mature in his *Ueber psychologische Methoden*, the leading article to the first volume of the *Philosophische Studien* (1883). Search ought to reveal something suitable on *Elementen* and their *Verbindungen*, the crucial topic in Wundt's system.

Perhaps some of *Physiol. Psychol.*, 1902, I, 339–350 would do, if nothing better turns up. (3) The only reason for quoting H. H. Donaldson on the temperature spots is that he is already in English. Blix or Goldscheider would be more appropriate, but I should quote von Frey who brought this skin theory together and defended it. If the articles in the 1890's do not yield suitable sections, then his *Vorlesungen über Physiologie* (1904) will. (4) Cattell has three papers from Wundt's laboratory (1885–1887) cited, besides the famous one on mental tests. It is not fair to the Leipsig laboratory to have Cattell its sole representative. Wundt's remark that Cattell was "ganz Amerikanisch" might be extended to include the functional orientation of Cattell's research. Nor does Cattell merit more quotation than any other author. I should keep one of his three, and add from Wundt's *aegis* two others which have special historical significance—perhaps L. Lange on reaction (beginning the experimental psychology of attitude, 1888) and Meumann on the time-sense (1892). (5) And then I should omit Burks on nature-nurture because I do not think her paper had great influence.

Those five changes would reduce the sixty-one items to sixty. Now here are forty-five more that seem to me to be of an historical importance as great as the items already included. Besides enlarging the book, they would also fill in lacunae and provide greater dynamic continuity for the genetic process being sampled.

Let me begin with Germany and Austria. It would be interesting to have something from the astronomer Bessel on the personal equation (1823). Purkinje's account of the phenomenon that bears his name (1825) is short, and the student will enjoy being shown that Purkinje observed 'twilight vision' at dawn. It would be well to have something from Herbart (1824) about the limen of consciousness and the inhibition of ideas. Then there would be Lotze and I should hope the *Medicinische Psychologie* (1852) would yield a section on local signs. Hering ought to come in somehow. Perhaps his nativism (1864) is more significant nowadays than his color theory (1874). Next I should hunt for a suitable account of brain localization in Fritsch and Hitzig (1870) and for proper comment by Brentano on the nature of act and intention (1874). Mach ought not to be left out for he leads straight into Gestalt psychology. His *Analyse der Empfindungen* (1886) would provide a section about the positivistic answer to dualism or on space sensations. There would hardly be room for both. The best thing to take from Stumpf's *Tonpsychologie* would be the description of tonal fusion (1890), but there is also his significant adoption of the psychology of function in 1907. We might have both of them. In the wordy, polemical G. E. Müller it is hard to find brief quotable material, but what about his psychophysical axioms and his hypothesizing of cortical gray (both 1896 and related)? I have already suggested Wundt on elements and their combinations, but perhaps we should also have Wundt on reaction times, since mental chronometry characterized the Leipsig laboratory. It will not do to omit Külpe entirely, and his 1892 criticism of the subtractive procedure in reaction times (especially if Ludwig Lange is included) would do for the early Külpe, and his 1912 paper on thought for the later Külpe. In the Würzburg School, we might hope for Watt on fractionation (1904) and Ach on determining tendency (1905). To round out Gestalt psychology, we ought to have, I think, von Ehrenfels on form-quality (1890), something from Schumann on perception in his papers of 1900–1904 (the Gestalt psychologists do not accept Schumann in their ancestry, but I think he is significant), and certainly an excerpt from Wertheimer, for which I nominate a selection from the 1921 paper in the *Psychologische Forschung*. That leaves out of my list Loeb on tropisms (1890, or better 1899) and Lipps on empathy (1903).

With abnormal psychology ruled out, the French become less important. There are already six French articles included. Gall on phrenology (ca. 1825) and Broca on the speech center (1861) might be added. It does not seem right to include Binet and exclude Ribot, but so many of Ribot's books were on mental disease. Could his *Psychologie de l'attention* (1889) furnish a significant section? And then what about Piéron? Does not France provide one living psychologist to rank with America's twelve?

My recommendations for Great Britain are these. Robert Whytt on the reflex (1751). Marshall Hall on the reflex (1832 or 1837). Thomas Reid on the powers of the mind (1785) as faculty psychology and leading into phrenology and factor analysis. Bain on association (1855), in addition to James Mill and instead of J. S. Mill on this topic. Spencer on evolutionary psychology (1870–1872) if the right passages can be found. James used Spencer's *Principles* as the only text in his course in physiological psychology. Ward on the subject-object relation, cognition and conation, from the 1886 *Britannica* article or the 1918 text. Romanes (1883) to show how careful he was about the anecdotal method. Lloyd Morgan (1894) for his canon. Spearman on g (1904). C. S. Myers ought to appear, but I do not know of anything truly significant for his role in British experimental psychology. Perhaps we must leave him out. Godfrey Thomson can certainly be included on overlapping abilities. I think of the 1916 paper in the *Brit. J. Psychol.*, because it had influence. And then there is McDougall on instincts (1908) and on purposive behavior (1923). Do we put Bartlett in? On reminiscence perhaps?

Now the Americans. There are none before Stanley Hall (1883) in Dennis' list. From Hall onward eighteen of the twenty-five psychologists are Americans. Since 1900 there are thirteen papers by native Americans and only five others. Perhaps Dennis has already looked after America well enough. I do, however, have some additions to suggest. James ought not to be limited to his theory of emotion but to have something that shows his position vis-à-vis functional psychology and Gestalt psychology. That can come from the *Principles* (1890). J. M. Baldwin should appear on evolutionary psychology but I do not know what to quote (1895 or 1897). Münsterberg should come in on applied psychology. Titchener, in addition to the paper on structural psychology, might be quoted on introspection (1910) and the nature of feeling (1908) or of thought (1909). Jennings on consciousness in protozoa (1904) is needed to offset the tropisticists. E. B. Holt on consciousness as reaction (1915) ought to be in, because he leads into Tolman who leads into operationism and modern behavioristics. Yerkes is now in on the Army intelligence tests, but I should like also to see him standing by his apes. Tolman should somehow receive the ball from Holt. And Thurstone should be there as the terminus of the series Reid, Gall, Galton, Cattell, Thorndike, Binet, Spearman, Terman, Thomson.

The net result of all this discussion is the suggestion to delete four of Dennis' sixty-one papers and to add forty-eight new ones, making a new total of 105.

Now let us see how this is related to countries and languages. We shall count contributions, not men.

	Germany and Austria	*Great Britain*	*America*	*France*	*Greece*	*Italy*	*Russia*	*Total*
Dennis	12	16	23	6	2	1	1	61
Dennis and Boring	34	29	28	10	2	1	1	105

Of the twenty-nine from Great Britain, nineteen were prior to 1883, which is the date of the earliest American paper. The associationists, the physicists and the biologists of Great Britain contributed most heavily to the early preparation for modern psychology.

The ratio 34/28 would not seem to me to weight the German papers too heavily in relation to the American. The ratio would be still larger if you stressed the nineteenth century more heavily in relation to the twentieth. Somewhere in the nineteenth century the Germans took over from the British, somewhere in the twentieth before Hitler the Americans passed the Germans.

Thus we arrive at the problem of translations. To be used as a text in American universities, such a book must be printed entirely in English. Dennis got along with the minimum of translation, for, of the twenty-two contributions in five foreign languages, nineteen were already printed elsewhere in English translation. The two pages of Mariotte on the blind spot and the eleven pages of Flourens on the nervous system were specially translated from the French for this book. Two pages of Helmholtz on the rate of the nerve impulse were translated specially from the German. How much new translating would be required if my proposal for a total of 105 contributions were to be put into effect? Twenty-four papers, twenty in German and four in French, with the possibility that two German translations could be saved if you took the less satisfactory but already translated books of Herbart and Lotze. That is a very considerable undertaking. It might run to three hundred pages of new translation and it is no wonder that Dennis did not undertake it.

Has not Dennis perhaps here made an answer to the contention of the Minnesota psychologists that the young new psychologists do not need to spend their time learning French and German because "in the few instances where a graduate student needs to consult a foreign language publication, it would be more economical of the student's time and energy for him to have the material translated for him" (*Amer. Psychologist*, 1947, **2**, p. 136)? If it were so simple as all that, Dennis could use the facilities which graduate students use at Minnesota and include my twenty-five German and French papers in English translation in his second edition. Certainly if we are to get those translations, Dennis must get the work done for us. When he has passed on, his successor may not know enough German to decide what to translate. You do not simply send the 53,735 German pages which Wundt wrote down to the university's translating bureau for an 'English' of them, so that you can then decide which are important enough to use.

About this useful book of Dennis' I have one minor bitterness, directed more at the published than at the editor. The right-hand page heading shows the topic being considered. The left repeats 'endlessly' (266 times) the title of the book. No chance that the reader will forget what book he is reading, and also but little chance that he can find what he is looking for by leafing through to it. I tested myself out guessing authors from the sixty-one page headings and missed on twenty. "Power of the Mind": that ought to be Reid, ought it not?—or Dugald Stewart, or maybe Gall. It *is* James Braid, of all persons.

All in all the book is a good job and Dennis earns our gratitude. The courses in the history of psychology will have an additional text, and the browsing graduate student with scholarly inclinations will often be found reading it.

EDWIN G. BORING.

Harvard University.

LEWIS, D. *Quantitative methods in psychology*. Iowa City: Don Lewis, 1948. Pp. 286. (Lithoprinted.)

This is the first book to present an account of quantitative procedures for psychologists at a more basic level than the usual treatise on statistics. In *Quantitative Methods* Lewis provides material for training the psychology student in basic mathematical techniques and gives him opportunity, in appropriates exercises, to apply the techniques to problems of psychological interest. If it happens, as a result of the approach exemplified by Lewis, that quantitative psychological theory is made meaningful to a large number of students and eventually becomes a matter of interest and practice to all well trained psychologists, then *Quantitative Methods*, as one of the first formal symbols of such a state of affairs, may have an important place in the development of psychological science.

Chapter 1 deals with some elementary principles and Chapter 2 considers the fitting of straight lines to empirical data. Chapter 3 is on logarithms. All of these chapters are straightforward and clear.

The use of non-linear curves is treated in Chapter 4. The chapter does not give an exhaustive account of the great variety of curves available for representing data, but the material covered is presented in a useful way. In only one respect do the reviewers disagree with the treatment of curve fitting: the application of a special case of the method of least squares to all forms of data. It is obvious that the method of least squares is one which is based upon certain assumptions, and if the assumptions are not fulfilled by the data, then the results of using the method are uninterpretable. When Lewis fits curves other than straight lines to data, he achieves the fit by transforming the data to yield a straight line and fitting this transformed function by least squares. Such a solution involves many assumptions, some of which are mentioned, none of which is emphasized. The problem of what deviations from best fit to minimize is always an important problem. If the data warrant elaborate treatment, the rationale of the steps used in fitting a curve to the data may be developed and may involve a complex analysis. It the data do not warrant such treatment, the fit may be chosen more obviously on a basis of convenience. In the latter case, some of the other methods mentioned by Lewis will yield easier and equally fruitful solutions. Considerations respecting least squares also apply to the discussion of the index of determination (p. 81).

Chapters 5 and 6 present a short account of calculus. Differentiation is treated in Chapter 5 and integration in Chapter 6. The treatment of calculus is probably less extensive than may be desirable. The discussions of differentiation, integration, and partial differentiation are condensed, and there is no consideration of methods of integration or of the use of integration tables. The problems of repeated integration and elementary differential equations are not mentioned.

Most of the latter part of the book (Chapters 7, 8, 10) is concerned with problems of the normal curve and problems of goodness of fit. Chapter 7, on the normal curve, is good and Chapter 10, on the fitting of straight lines for testing the significance of parameters is also valuable.

Some portions of the treatment of distribution functions (Chapter 8) and goodness of fit (Chapter 10) will be much less obvious to students than the discussion of earlier topics, such as logarithms and differentiation, and the use of the book as a text may require considerable supplementary lecturing. There may be some question, for example, as to the value of presenting the formula for the chi-square distribution (p. 174) without showing how one arrives at it; Lewis uses the formula to arrive at the mean and modal values of such a distribution. Such a procedure may help to eliminate some fear of large formulae but, from the point of view of understanding the distribution, the student might just as well be given the mean and modal values arbitrarily.

Chapter 9 is one of the most valuable chapters in the book. It gives the student an opportunity to apply the principles he has learned to a connected background of theory in his own field. The treatment of Hull's theory (p. 233 ff.) is good, and the description of Hecht's photoreceptor hypothesis is valuable, although presentation of a later version of the theory would be desirable. Finally the reviewers believe that Woodrow's equation (pp. 247 ff.) receives more extensive treatment than it merits.

One is tempted to compare *Quantitative Methods* with books in other fields written to serve comparable purposes. In physics, chemistry, and engineering, one encounters treatises on quantitative methodology which range in difficulty from Daniels' *Mathematical Preparation for Physical Chemistry* to Mellor's *Higher Mathematics for Students of Chemistry and Physics* and the Sokolnikoffs' *Higher Mathematics for Engineers and Physicists*. Lewis' book lies in difficulty somewhere between the level represented by Daniels and levels represented by Mellor and the Sokolnikoffs. Despite the fact the Daniels book is written at a more elementary level than Lewis', it seems to the reviewers that the Daniels work is more satisfactory from the point of view of training in basic mathematical techniques, particularly in applications of calculus.

It may be argued that a course which is primarily of psychological interest should not treat topics such as calculus which can be learned in other courses. Nevertheless, it has been the experience of the reviewers that a course in quantitative treatment which emphasizes the application of calculus is of great benefit in elucidating the concrete problems in a student's own field.

A great deal of space in *Quantitative Methods* is taken up by statistical theory. This orientation is an important one and must not be underestimated. Nevertheless, it presents certain difficulties, not the least of which may be the development of a type of thinking which is more concerned with establishing differences than with describing functions. An additional danger lies in the fact that the student may become submerged in details of special statistical techniques without receiving an understanding of mathematical thinking. The reviewers agree with Lewis that psychological theory can benefit greatly by the application of mathematical thinking to many problems. However, they have a prejudice that basic mathematical techniques rather than statistical analysis should be emphasized in an orientation course on quantitative treatment. Training in advanced statistics may progress more effectively in later courses against a background of well learned mathematical operations.

Despite our opinion that Lewis's account should be strengthened in its treatment of basic topics, the book has much to recommend it. Let there be no question about the attitude of the reviewers: Lewis has written a valuable book. It is the first book to present an appreciative account of materials requisite to the quantitative thinking of psychologists, and, for this reason, it should receive wide attention and influence the graduate training of psychologists in a desirable way. If the book stimulates the mathematical training of psychologists, it will be a major contribution, and other books dealing with the same topic may be expected to follow. If these books improve on Lewis', it is likely they will do so by emphasizing the rationale of analysis rather than the techniques. In any case, *Quantitative Methods* evidences an appreciation of an important gap in the training of psychologists and fills the gap with no little success.

C. H. GRAHAM.
C. G. MUELLER.

Columbia University.

RUCH, FLOYD L. *Psychology and life.* (3rd Ed.) New York: Scott, Foresman, 1948. Pp. 740.

Ruch's new edition of *Psychology and Life* is an enlarged, better written and illustrated version of his earlier volumes.

In evaluating such a textbook one must of course acknowledge that the merit of a particular book depends upon the instructor's interests, bias, and philosophy of the purpose of psychology. Relative merit also depends upon the level of ability of the student, the major course the student is pursuing, and the probability of his taking further psychology courses.

If the instructor wishes to emphasize the physiological-experimental-scientific approach, Ruch's third edition probably falls short of Munn's *Psychology.* If the instructor wishes to remain in the traditional frame of perception, sensation, learning, memory,thinking, but with perhaps a bow to the application of psychology to life, the *Psychology* of Woodworth and Marquis is probably more suitable than *Psychology and Life.*

However, for the instructor whose students will have little further work in psychology and who expect some immediate answers to everyday problems, Ruch's revision has no competitor. Whereas, his former editions have been overly talkative in parts, have admittedly stressed subjects "interesting" to the student, have been weak in systematic presentation, the present edition has none of these weaknesses. Statements are well documented. The writing is direct, easier to understand, better organized. In the trend of picturization so well done by Munn, Ruch has added a variety of excellent drawings, some in color, and a number of well chosen photographs.

Considerable material has been added by reducing the size of print and tightening the exposition. Several chapters are now the result of collaboration with others: two chapters on learning (with Robert Leeper); "Emotional Development: Clinical Findings" (with Joseph Shoben); and "Illustrated Reference Manual: The Brain and Nervous System" (T. C. Ruch). Several editorial consultants are acknowledged.

A section on emotional adjustment includes some 135 pages, and is a distinct contribution for the student who is looking to psychology to give him some insight into his own and others' behavior. Techniques of psychotherapy such as non-directive counseling, finger-painting, play-therapy, and psychodrama are presented in an interesting and basic manner. A new chapter on "Problems in Human Relations" which discusses marriage, parenthood and delinquency is characteristic of the movement into the applied area.

As to adverse criticisms, the revision will appear to many to but emphasize an already undesirable "popularizing" of psychology. It can be said for example, that it is unsound to introduce the topic of marriage in a beginning course in psychology, that marriage is the province of sociology, not psychology. Others who may accept marriage as a field for psychology will affirm that nothing good can come from a seven-page treatment. Many will be querulous about including such paragraphs as "What About Dunninger?" and "Is there Extra Sensory Perception?" (Students ask these questions even if the textbook writer does not.) Again, as in previous editions, the good rats do not run the students through mazes, although more attention is given to them than previously. This will no doubt continue to be construed by a few as an undesirable movement toward understanding people.

In fine, Ruch has continued to popularize his writing—but most of the previous criticism concerning looseness, lack of fact, and lack of a basic approach, have been met. Ruch has played to his audience, as any good teacher must—and why not?

The third edition of *Psychology and Life* is a real contribution to understanding life through psychology.

DONALD E. LUNDBERG.

Cornell University.

DAVIS, ROBERT A. *Educational psychology*. New York: McGraw-Hill, 1948. Pp. x+349.

This book is divided into two main parts. The first, entitled "Understanding the Learner," includes chapters on "Physical Growth and Health," "Scholastic Ability," "Interests," "Attitudes," and "Emotional and Social Maturity." The second, entitled "Directing Learning," takes up in successive chapters "Preparing Materials for Learning," "Cultivating Abilities," "Testing for Learning," "The Gains of Learning," "Maintaining the Gains of Learning," "Favorable Conditions for Learning," and "Incentives."

In an informative preface the author states that the book was planned for use in introductory courses in educational psychology and was intended to provide information concerning important problems of the classroom and the school, problems which teachers meet in their day-to-day work. Although psychologists will inevitably differ to some extent regarding the subject matter that teachers will find most useful, I think most psychologists will agree that the content of this book has been well selected. What is equally important, I think there will be wide agreement that the book is unusually well written. The author's style is lucid and smooth.

The unique feature of his book is that the traditional subdivisions of educational psychology have been almost entirely ignored. One will look in vain for, say, a discussion of the law of effect or of Gestalt psychology. For this reason, the book nay not appeal to some who organize and teach their courses in educational psychology in the traditional pattern. It seems to me, however, that good reasons can be adduced for reorganizing beginning courses in educational psychology to make them more functional, more interesting, and more nearly up-to-date. Dr. Davis has provided not only an interesting and well written text but a stimulus to further experimentation in presenting the subject matter of educational psychology.

FREDERICK B. DAVIS.

George Peabody College for Teachers.

GHISELLI, E. E., AND BROWN, C. W. *Personnel and industrial psychology.* New York: McGraw-Hill, 1948. Pp. xiii+475.

As a survey of the major applications of psychological techniques to personnel problems in business and industry, this book is designed to meet the needs of university students, as well as those of persons in business and industry who desire to obtain a foundation in the basic principles of industrial psychology. The book represents a culmination of four earlier mimeographed editions, which were used in classes at the University of California over a period of several years.

The sixteen chapters are arranged in a logical sequence. Following the introductory chapter, in which emphasis is placed upon the scientific method as applied to *human* (psychological) as contrasted with *mechanical* (physical) factors in production, the topics of job and worker analyses, criteria of job proficiency, and rating methods are successively treated. An understanding of these techniques is basic to all of the topics subsequently presented. The next chapter deals with the interview and personal-data analysis, and is followed by three chapters dealing with various aspects of the validation and use of psychological tests. Subsequent chapters treat the definition and meaning of effective work (including efficiency, fatigue, etc.), conditions of work and productivity (diurnal changes, length of work period, rest pauses, shift work, physical conditions of the work environment), effective methods of work and the design of tools and equipment (time and motion study, work simplification, layout, and what has recently come to be called "bio-mechanics"), industrial training (two chapters), accidents and safety, monotony, and worker motivation and morale. The volume contains 115 tables and 38 figures (all line drawings), selected references at the end of each chapter, and terminal indexes by author and subject.

The general format of the book is identical to that of other volumes in the McGraw-Hill series, under the Consulting Editorship of Dashiell.

In the reviewer's opinion, this book provides an excellent presentation of the basic facts and methods of industrial psychology, and represents the best textbook currently available in this field. The presentation is sound and lucid, and there is repeated emphasis upon the need for empirical validation of techniques. Although classical material is well represented, frequent reference is made to recent experimental data, such as those obtained in World War II. The *limitations* as well as the *possibilities* of various techniques are clearly indicated. Thus, for example, in the discussion of the limitations of personality questionnaires it is excellently stated that the validity of such a questionnaire "will be a direct function of the honesty and accuracy with which the individual answers it" (p. 169). Hence the reader is prepared for the conclusion that personality inventories are far more valid in guidance or counseling than in selection situations.

Despite its many virtues, however, this book shows a few unfortunate short-comings, many of which represent errors of *omission*. Certain topics, such as job evaluation, are not treated at all, although such an omission is, curiously enough, characteristic of most textbooks by psychologists in the industrial field. Nor is any reference made to market survey techniques—an omission which is less serious since the authors evidently regard consumer research as outside the field of industrial psychology. Another minor but important omission occurs repeatedly in the tabular material throughout the book: experimental data are often reported without indication of the size of the sample on which such data are based; nor does the accompanying text reveal the number of cases in such instances.

Incentive systems are also inadequately treated. Although rating methods and their pitfalls are thoroughly discussed, little is said regarding the importance of *training* the raters. The distinction between *production* and *productivity* is not clearly drawn (cf. e.g., pp. 12–13, 239)—a distinction which is currently recognized as of fundamental importance in the labor relations field. Throughout the book, furthermore, the authors emphasize the importance of applying scientific techniques to the *worker*, but the application of similar techniques to all levels of *management* is either ignored or under emphasized, despite their equal relevancy in both fields of application. Moreover, the authors' contention that "the facts point to the 'sore spots' of production lying with the worker" (p. 17) is somewhat at variance with the opinion of many consultants as well as with the authors' own statement that the key to "more effective work through better worker morale" is "more understanding supervision" (pp. 557–458). Certainly, most industrial consultants would emphasize the critical rôle of at least "middle management" (foremen and supervisors) in the current industrial picture.

In the introductory chapter and in certain other places throughout the book the authors state a theoretical point of view which may be misinterpreted by those psychologists who are quick to applaud any deprecation of objective and rigorous formulation of psychological principles. Thus, the authors state that "interest should be centered upon the worker as a thinking, feeling, and desiring organism, and *not* as a mechanical device that responds in given ways to appropriate stimuli" (p. 2). The intent of the authors would appear to be that of emphasizing the *error of a priori assumptions* regarding the worker's motivational stimulation, the importance of considering the *motivation and attitudes of the individual worker*, and the need for experimental investigation of such motivation and attitudes in *specific working situations.* Unfortunately, however, many readers with a mentalistic bias will misinterpret such a statement, and will construe it as a denunciation of the applicability of S-R concepts—or even of the scientific approach—to the industrial field! The writer feels certain that Ghiselli and Brown would not subscribe to such an obscurantistic viewpoint.

One further point deserves special mention. In reviewing the historical background of the applications of psychological and engineering techniques to production problems (cf. pp. 6–7, 14–16), the authors indict management—and indirectly the engineer and the industrial psychologist—for having in the past directed attention solely to the goal of increasing production and hence of increasing profits. It is pointed out, to be sure, that industrial psychologists have, in recent years, fortunately turned their attention to problems much broader than that of "production and profits." The authors fail to emphasize, however, that *the services of competent consulting psychologists are available to labor as well as to management*, and that *the failure of labor unions to utilize such services cannot be laid at the door of either management or the consulting psychologist.* Although this point is not made by the authors, there is a hint of it in a later section of the introductory chapter. In discussing the need for research under actual working, rather than artificial laboratory, conditions, the authors state that such an objective "would require that industrial and union organizations jointly provide the means for establishing research groups within the industrial organization" (p. 18).

Everything considered, however, the merits of this book clearly outweigh its shortcomings. It represents a real contribution to a field in which there are all too few good modern textbooks.

John P. Foley, Jr.

Industrial Division, The Psychological Corporation.

Burtt, Harold E. *Applied psychology.* New York: Prentice-Hall, 1948. Pp. x+821.

The advent of a new textbook whose author is a recognized pioneer in his field is always an interesting event. Most of us are engaged in the never-ending task of finding a text which best suits our needs and efforts. More important, however, is the fact that such a book inevitably implies a definition, or at least a delineation, of a field. These definitions, more frequently than those of the more concise and formal type, may affect materially the subsequent course of work in the defined area. This is particularly true if, as in this case, the author writes for students and laymen (perhaps actual or potential employers of applied psychologists) alike.

It must be recognized, in short, that Dr. Burtt's book will have considerable influence upon what psychology instructors, their students, and laymen think applied psychologists should do and how they should do it. Its evaluation in this respect, as well as a pedagogical instrument, therefore seems in order.

The scope of this volume is wide. Apparently, the author conceives of applied psychology as a field which includes all efforts made by psychologists or other individuals to help anyone, regardless of whether the principles or methods of scientific psychology are involved. After an introduction and a quick tour of the field of general psychology, there follows a series of chapters which inspire the reader with considerable respect for Dr. Burtt's breadth of interest and mastery of a voluminous literature. Three chapters are devoted to the contribution of psychology to education, two to personal efficiency and vocational guidance, three to medicine, four to law and crime, two to employment, three to industrial efficiency and morale, four to advertising, and one to "outlying fields" which include "Life Enjoyment" (the arts, recreation, and athletics), "Communications" (public speaking, writing, drama, movies, radio, television, and the press), and "Social Progress" (public opinion, politics, religion, war, and peace).

This implied definition of applied psychology will probably be as disturbing to some readers as it is to this reviewer. These readers will wonder if perhaps the field does not need a re-definition which would make it unnecessary for the author of a textbook to include a treatise on the nervous system, psychiatry, the problems of social workers, or the practices of advertising men. They may believe that it is more important for the student of applied psychology to understand the basic logic and techniques which are applied in the various fields than it is that he be familiar with the reflex arc concept or the alleged effects of frontal lobotomy. They may even feel that it is time applied psychology *defined* its field rather than embracing whatever can be engulfed by it.

These readers will soon find that Dr. Burtt is in disagreement with their position. His emphasis is upon results—so much so that methodology is given very short shrift. The consequence is a series of deficiencies and inconsistencies that will distress the trained reader but will probably (and unfortunately) go unrecognized by the neophyte.

For example, the problem of sampling, so basic and so deceptive to workers in the applied field, is apportioned one page and this, many ful attention to detailed procedures has, in certain instances, resulted in repetition or reduncancy, although this tendency has not unduly marred the effectiveness of the book.

The discussion of procedures is distinctly not of the ivory-tower variety, since recognition is constantly given to those practical situations which may preclude a close hewing to a given theoretical line. For example, the authors make no bones about the acceding to the interests of either management or labor on the inclusion or exclusion of any given factor, in order to keep peace in the family.

The most vulnerable spot in the book seems to be the acceptance of the point of view that the selection and weighting of factors must of necessity be made on the basis of the best "guesses" (the authors use the word "judgments") of those responsible for the development of the system, relying on the past experience of other companies as a guide. It is entirely feasible for a company to determine factually the relative importance of the factors that are used in its system; the statement that "there is no known method and no known criteria for determining the relative values in an objective manner" seems just a bit strong in the light of the fact that such relative values *have* been determined in specific situations.

As a contribution to clear thinking in the field of job evaluation the book is of material significance. With full awareness of psychologists' taboos about dealing with superlatives, it is the opinion of this reviewer that the book is the most adequate over-all manual on the philosophy and techniques of job evaluation that has yet appeared.

C. H. LAWSHE.

Purdue University.

KAPLEN, OSCAR J. (ED.) *Encyclopedia of vocational guidance.* New York: Philosophical Library, 1948. Vol. I, pp. xxi+722. Vol. II, pp. 700. $18.50.

Three hundred authorities in eleven countries contributed articles to this *Encyclopedia.* Topics which are covered include: vocational guidance activities throughout the world; vocational guidance periodicals and organizations; the handicapped; problems of minority groups; aptitude patterns in the principal occupations; occupational trends in the United States; economic aspects of guidance; techniques of interviewing: evaluation of guidance procedures; administrative aspects of vocational guidance; descriptive articles on the principal psychological tests; and other subjects.

The editor states that this book "is encyclopedic only in a relative sense; it does not pretend to cover in great detail all phases of the guidance field. The topics included were chosen from a master list and many important subjects were omitted because of lack of space." The encyclopedia is designed primarily for professional counselors who do not have large library resources.

Because economic aspects of vocational choice are so important, articles have been prepared on occupational trends, unemployment cost of preparation for the major professions, use of census data, and average earnings. No attempt has been made to present information on the training requirements, economic outlook, or general nature of individual occupations.

Topics are arranged alphabetically in formal encyclopedia style. For example, in the first 15 pages we have the following articles: "Accountancy, Aptitude For"; "Achievement Tests, General"; "Achievement Tests, Use of"; "Acrophobia"; "Alliance for Guidance of Rural Youth"; "American Council on Education Psychological Examination"; "Acting, Aptitude for."

It must be emphasized that these two volumes comprise reference material and are to be used for the purpose of acquiring pertinent information quickly. It should be used in the same manner that a counselor uses the *Dictionary of Occupational Titles*—when a need arises. For example, if a counselor is called on to assist a blind person or a hard of hearing individual or a member of a minority group, a quick reading of the pertinent topic beforehand will be of great value.

Many of the articles are descriptions of principal psychological tests. It is difficult to avoid comparing these presentations with the descriptions and critical evaluations found in Buros's *Mental Measurements Yearbook*. The discriminating counselor will, I am sure, find the *Mental Measurements Yearbook* a more helpful tool in deciding what tests to order or to use. Evidently, Dr. Kaplan has a conflict about including material on tests. An "Encyclopedia of Vocational Guidance" could not well omit such information; on the other hand, the *Mental Measurements Yearbook* had both description and critical comments. Therefore, the editor took a compromise stand:

> The primary purpose of the test articles has been that of description, although critical evaluations also appear with regard to the principal categories. The discriminating reader is referred to the *Mental Measurements Yearbook* edited by O. K. Buros (p. vi).

If one has the *Mental Measurements Yearbook* in his library—everyone using tests should—the articles on tests in this Encylopedia are of little value.

Articles of great merit are the critical evaluations of classes of tests, such as, "Art Tests, Critical Evaluation of the Principle" and "Clerical Aptitude Tests, Evaluation of." Many critical comments on specific tests will be found in these summaries. From the user's standpoint it means reading at least two articles to find complete information on a given test.

The physical make-up of the books causes annoyance. There is no indication that Vol. I contains topics A through L and that volume II starts with M. Within a volume it is difficult to find a given title because of the lack of any indexing. Appropriate letters on the corners of the pages or thumb-nail indexing like a dictionary would facilitate quick reference.

Despite the objections mentioned the Encyclopedia is a contribution to the field of guidance. Along with the *Dictionary of Occupational Titles* and Buros's *Mental Measurements Yearbook* it should have its place on the bookshelf of every vocational counselor.

STANLEY G. DULSKY.

Chicago Psychological Institute.

NAUMBURG, MARGARET. *Studies of "free" art expression of behavior problem children and adolescents as a means of diagnosis and therapy.* New York: Nervous and Mental Disease Monographs, 1947, No. 71. Pp. xi+225.

This monograph is a collection of six articles published by the author in various psychiatric journals and in a volume on *Child Psychiatry* during the period between April, 1944 and January, 1946. All six studies are part of a research project at the New York State Psychiatric Institute and Hospital and are devoted to "the investigation of the possible use of 'free' or spontaneous art expression as an aid in both diagnosis and therapy" (iii). Each study demonstrates the extent and nature of the fulfillment of the project's aim by giving a detailed analysis of one child's graphic productions, conversations and general behavior during his periodic art sessions with the author.

By means of numerous reproductions Miss Naumburg shows convincingly how impersonal routine tracing and copying, common among public school children, gradually is replaced by spontaneous expression of personally significant subject matter: "As soon as original art work in encouraged, instead of dependence on models and specific techniques, the focus of a patient's art activity is modified. He will begin to draw on his own inner resources and this will inevitably lead to some expression of the conflicts within the personality, which may reveal aspects of the pattern of his mental disease as well as the specific insecurities or traumatic experiences within the patient" (p. 50). By releasing these largely hidden forces the author feels that the art productions of her patients serve the dual purpose of aiding the psychiatrist in diagnosis and in therapy. Whereas in a setting of this type both functions are more or less interwoven, the former goal is served primarily by bringing to the fore hitherto unverbalized diagnostic material, thus corroborating, anticipating or broadening psychiatric findings. The therapeutic contribution is in form of catharsis, substitute satisfaction for suppressed desires and strengthening of self-esteem through artistic accomplishments.

Although the author's main purpose is to demonstrate the value of free art expression in clinical practice, occasionally she reaches out beyond her clinical material by calling imaginative expression "a source of growth and sustenance as well as a language of communication in the life of every individual, whether he happens to be mentally disturbed or at peace with himself" (p. 89). In fact, in one place we find the far-reaching assertion: "In the East unstinting recognition has always been accorded to art as an expression of and not as an escape from reality; could the West bring itself to an acceptance of all forms of creative expression as a universal normal and integrating experience that is neither effeminate not neurotic, our culture might again find ways to restore harmony and balance to the disequilibrium of the modern psyche" (p. 90). Whether or not we follow Miss Naumburg all the way in these contentions, it cannot be denied that creative activity has a cathartic as well as constructively fulfilling effect. Any efforts to encourage its broader use in such fields as education, recreation or therapy can therefore only be welcomed.

So much for the main theme of this monograph. The clinically interested reader will surely find it stimulating and worth looking into. Those who wish to try out the technique discussed in it will appreciate both the specific suggestions regarding the author's actual procedure of starting the process of "liberating through art" as well as illustrations of her use of verbalizations concurrent with the graphic expressions. A person with wider interest, however, might well experience a certain amount of disappointment after finishing this book. One could wish that some attempt had been made to realize the potentialities of this rich and suggestive material for an advancement of our understanding of general theoretical questions such as physiognomic quality in perception, the psychological nature of symbolism, etc. Particularly the last study in the volume should lend itself well to theoretical exploration. It should be remembered, however, that the aim of Miss Naumburg's study was clearly stated as practical and not theoretical in nature and we all know how difficult it is to do equal justice to these frequently divergent aims within the framework of the same research project. The theoretical potentialities of this material are primarily suggested, therefore, as a realm for further expansion.

Finally it is regrettable that this monograph bears evidence of hasty preparation for the press. The articles, which had already been published separately, have been here thrown together into book form without any attempt at unification or organization except for a brief preface by the author and a foreword by Nolan Lewis. The same or very similar generalizations are repeated in many places. There is even one instance where a long passage from one article is quoted verbatim in another, when a mere page reference would have sufficed. In addition to this repetitiousness among the separate articles, lengthy descriptions of the reproduced pictures appear in practically identical form in both the text and the caption of the picture. In short, greater editorial thoroughness would have much enhanced the readability of this worthwhile report.

MARIA A. RICKERS-OVSIANKINA.

Wheaton College.

LINNDER, ROBERT M., AND SELIGER, ROBERT V. (EDS.) *Handbook of correctional psychology.* New York: Philosophical Library, 1947. Pp. 691.

The kindest thing that can be said about this volume is that is a pioneering venture in an area that has been very inadequately explored. Although the authors titled the book "correctional psychology," and the inference is that by "correctional" is meant therapy in the penal sense, an analysis of the articles reveals less than one-quarter would not fit primarily into pure medicine, physiology in its applied phases, or sociology, especially in its forensic aspects.

For example, the reviewer was surprised to find chapters such as: "Electroencephalography: used in penologic practice"; "Sugar metabolism in its relations to criminology"; "The intramural practice of eye, ear, nose and throat"; "The acute medical patient"; " A venereal disease program in prison"; "Use of sedative measures in mental hospitals"; "Treatment of traffic offenders." While each and every one of these and other chapters can be argued for in a volume on criminal psychology, to say that the sum of articles such as these are "psychology" is to stretch the semantics a bit!

The bug-a-boo of all editors of compilations—variations in style and quality is well evident here. Some articles, like Brancale's "The classification clinic in a correctional institution" and Selling's "The extra-institutional treatment of sex offenders" are excellent, appear to belong logically within the area of discussion, and make the book valuable for those in the field, while other articles seem either so peurile, forced, or down-right silly that they make the reader wonder why they are included in a serious volume apparently designed for the use of specialists.

To whom is the volume addressed? It is too specialized to be used as a text in criminology; it is too superficial for the use of specialists in any area in penology; it is the source of too many original articles to be passed by. It is probably best suitable for a text for people in in-service courses in prisons who need to know the functions and problems of associated disciplines.

The need for a book that will discuss the origins, dynamics, therapies for criminal behavior has not been met. The present volume is not wellbalanced, but then neither is the field of penology these days. The fact that the editors were unable to find a single member of the American Psychological Association who was employed in the penal area to enter into this collaboration of 46 minds is a sad commentary on the present state of penal psychology.

The atrocious technical aspects of this volume deserve comment. The very first word of the book is mis-spelled—Editor's instead of Editors'. We learn there is a test known as "Revised Standard—Binet." In the Table of "contents" the titles of the chapters are not linked with the names of the authors. Some chapters have bibliographies, some do not. One bibliography is about 300 pages out of place. There is no index. The articles are not sectioned off in any logical manner. For the price, a more substantial binding would have been *en regle*.

In sum, it would be fair to say that there are various chapters that are well written and are unique and for these chapters the book is well worthwhile; and that while the assembly job seems sloppy that there is enough gold in the volume to recommend it for the person interested in correctional psychology.

R. J. Corsini.

San Quentin.

Cottrell, Leonard S., Jr., and Eberhart, Sylvia. *American opinion on world affairs in the atomic age*. Princeton: Princeton Univ. Press, 1948. Pp. xxi+152.

One of the greatest scientific achievements, and perhaps the most powerful psychological weapon of World War II, was the atomic bomb which was dropped on Hiroshima on August 6, 1945. This and the later one dropped on Nagasaki were primarily responsible for bringing the war to an end, and opened a new era in man's understanding of warfare. As one leading physicist has pointed out, it was not one atomic bomb, or two, which brought surrender, but rather the experience of what one atomic bomb will do, *plus the dread of many more* that was effective.

Atomic scientists have pointed out that nuclear knowledge of this kind cannot forever be kept secret, and that no adequate defenses are known or can be devised. It is also conceded that the use of atom bombs in another war could mean the end of civilization as we know it. Because of the enormity of the implications of these new scientific developments, plans were formulated immediately after the war for the development and control of atomic energy for *peaceful* uses. One of the tests made on the bomb after the war, known as "Operation Crossroads," may well have been significantly named—man's choice of the use of atomic energy for peace or war. On a purely rational basis, the decision would seem clear.

As a psychological factor, the atom bomb had an indisputable effect on the whole world. All the old concepts of war were changed overnight, and a covert fear gripped people of all nations, including those of the United States, which was at that time the only country in possession of atom bombs, or of the "secret" of producing them. The existence of the bomb imposes a great moral obligation on people throughout the world to outlaw its use for war; and the means of its ultimate control rests with a strong international organization.

This is certainly the view of most physicists and atomic scientists, and it was not long after the war before organizations were formed such as the Emergency Committee of Atomic Scientists, and the Association of Education for Atomic Energy.

The present book by two social scientists is based on a report prepared for the Committee on Social and Economic Aspects of Atomic Energy of the Social Science Research Council. It is primarily a summary of a study made in 1946 of American attitudes regarding foreign affairs, especially in relation to atomic energy.

The most interesting thing psychologically is not simply the vast ignorance of large segments of our population regarding world affairs—for example, one-third of the people interviewed could not give even the simplest answer to the question, "What is the main thing the United Nations organization is set up to do?"—but rather the contradictions in attitudes. For example, the majority of people interviewed indicated that they approved the idea of going beyond the idea of a United Nations to a world government, but at the same time objected to lowering tariff barriers. Likewise, most people were veering away from the notion of isolationism, but at the same time did not think it wise to give or lend money to people of other nations. Of course, contradictory attitudes of which these are illustrations are well known to social scientists and others who have carried out attitude studies.

Regarding the atom bomb, most people apparently felt so completely helpless regarding its potential damage that in one sense it had almost ceased to have reality for them. Their attitude was largely one of "let the government worry." In fact, this same attitude of preferring to leave the entire problem up to the "leaders," the "experts," or the "government" has been illustrated in recent studies reported by the Society for the Psychological Study of Social Issues. The main lesson in all of this is the enormous psychological distance between the citizens of our democracy and the world issues that concern their elected and appointed leaders.

The present volume is presented in fairly technical form, including a great deal of tabular material. Apparently it was not written for popular consumption. An index at the end would have been heplful.

STEUART HENDERSON BRITT.

McCann-Erickson, Inc., New York City.

WOLBERG, L. R. *Medical hypnosis.* New York: Grune & Stratton, 1948. Vol. I, pp. xi+449; Vol. II, pp. vii+513.

These two volumes are concerned with the principles and practice of hypnotherapy. While the title of this work suggests material from the general field of medicine, its contents indicate that the author's main concern is with the field of psychopathology. It may be stated at the outset that Wolberg does not consider hypnosis as the ultimate in therapy, but rather as a useful and legitimate tool, which like any other therapeutic procedure, has its indications and contraindications.

Volume I includes brief but adequate discussions of the traditional topics in hypnosis, i.e., history, phenomena, tests, technics, theories. The greater part of the first volume deals with the direct and adjuvant use of hypnosis in such varied disorders as alcoholism, psychosis, conversion hysteria, anxiety neurosis and psychosomatic conditions. In addition there is a brief statement of the possible use of hypnosis in such miscellaneous conditions as overweight, insomnia, nail biting, stage fright and excessive smoking. Preceding the discussion of therapy in each of the above conditions is an analysis of the personality dynamics involved. Direct and complete recording of patient-doctor conversation in three cases makes up the bulk of the second volume. The shortest of these histories comprises 85 pages. The cases illustrate the use of hypnosis in symptom removal, psychobiologic therapy (guidance, reassurance, desensitization, reeducation) and psychoanalytic therapy, respectively. The use of hypnosis in any but the latter case involves, Wolberg warns, the risk of discharging a patient symptom-free but without a full understanding of his inner conflicts. Justification, therefore, for the use of hypnosis in symptom removal or psychobiologic therapy is to be found only by a careful scrutiny of certain aspects of the personality (ego strength, the role of the symptoms in the total personality, motivation to be symptom free), certain situational factors (money and time available), and the therapeutic goal considered feasible.

The question of therapeutic goals is too often ignored and it is here that Wolberg, by his thorough discussion of this and the related topic of symptom treatment, makes his most pertinent contribution. Also of interest is the author's theory of hypnosis and his observations concerning the relative effectiveness of hypnosis and narcosynthesis. In addition, both volumes contain detailed procedural suggestions relating to technics of trance induction, deepening of the trance, eliciting of hypnotic phenomena and self hypnosis.

The chief criticism of this reviewer concerns the tendency of the author to explain away the discrepant data concerning such problems as changed sensory limen, individual differences in susceptibility, and validity of regression in terms of aptitudes. These aptitudes are said to be determined by the interplay and balance of both conscious and unconscious strivings within the individual. The subject who is hypnotized "against his will" or by a disguise technique is thought to be strongly motivated toward susceptibility on the unconscious level, while in the cooperative but nonsusceptible person, the reverse situation would hold. The data supporting this hypothesis, however, are of a post hoc nature and the hypothesis itself is not easily subjected to experimental verification. Another objection is Wolberg's use of Rorschach test results in support of the reality of regression. This would seem to assume the existence of well standardized age norms.

The final chapter on the future of hypnosis is both brief and weak. The author argues that in the interests of both medical and lay groups, safeguards (i.e. laws) are required which would restrict the use of hypnotherapy to members of the medical profession. There are many objections to this conclusion. Not the least important of these is the fact that, as the author's own references indicate, many individuals not privileged to belong to the ranks of the medical profession have made contributions both directly and indirectly to the field of hypnotherapy.

In these two volumes, Wolberg has presented a well written and interesting work which makes a definite contribution to the field of hypnosis. These books may be profitably used as reading material in advanced courses in abnormal and clinical psychology.

F. L. MARCUSE.

Cornell University.

HALL, VICTOR E. (ED.). *Annual review of physiology.* Stanford, Calif.: Annual Reviews, Inc., 1948. Pp. xi+552.

For the psychologist not intimately concerned with specific physiological processes, there is relatively little of value in the volume. Of the 23 chapters, only five are oriented toward problems with which the profession commonly deals. These five are: (1) Vision, by A. Chapanis, 24 pages; (2) Physiological Psychology, by W. H. Gantt, 26 pages; (3) Heat and Cold, by D. H. K. Lee, 22 pages; (44) Anoxia in Aviation, by L. F. Nims, 10 pages; (5) Somatic Functions of the Central Nervous System, by W. S. McCulloch and J. Y. Lettvin, 16 pages. The remaining chapters contain isolated, relevant studies for the psychologist, but they are scarce.

Gantt's contribution on physiological psychology is frankly limited by the author to the consideration of his own field of interest, Pavlovian methodology and human psychopathology. However, for those similarly occupied, the references to current Russian experimentation and the expostulation of the need for a "level of integration" approach to psychobiology are rewarding. Psychologists need the fertilization which comes from such considerations of integrative levels as are presented in Novikov's paper in *Science* (March 2, 1945, **101**, 209–215). The reviewer found Chapanis' chapter a concise and worthwhile survey of the salient new studies on color vision. Considering the mass of material emerging from the war, it is an achievement to have isolated the significant studies as Chapanis has done.

McCulloch's review is concerned mainly with the research done in his laboratory and related areas. Most of the references cited at the end of the chapter are not described in the text. For those dealing with the physiological substrata of oxygen deprivation, Nims' brief summary will be of particular interest. It would have been valuable to have related the high altitude effects of anoxia with those reported for such psychosomatic therapies as insulin and electroshock. Lee's contribution is an effective presentation of the environmental pressures arising from exposure to heat and cold. Many of the data offered are rarely cited in psychological literature.

It is to be hoped that the editors of the forthcoming first "Annual Review of Psychology" will follow the example of the editors of this volume in their attempt to ensure author representation from all parts of the world regardless of "curtain effects."

BERNARD F. RIESS.

Hunter College.

MOORE, THOMAS VERNER. *The driving forces of human nature and their adjustment: an introduction to the psychology and psychopathology of emotional behavior and volitional control.* New York: Grune & Stratton, 1948. Pp. v+456. $6.50.

The reviewer hazards the guess that the late Father Moore of the Catholic University of America was the sole contemporary psychologist who was not only a Ph.D. and an M.D., but also (as the title-page declares) a "monk of the order of St. Benedict." Since stereotypes victimize those who "know about" them as well as those ignorant of the concept, some professional colleagues may be wrongly disposed to view this volume as just another re-statement of Thomistic doctrine. Actually, it is an original and powerful synthesis of experimental psychology and clinical psychiatry with "the Perennial Philosophy."

There are seven parts to this treatise which is pitched on an intermediate or advanced level:

The *first* is a brief personalized history of the foci of American psychology with one of the most penetrating treatments in print of the deficiencies in the logic and philosophy employed by William James. Moore's own agreeable modernity is revealed in his evident distaste for "the Latin terminology of certain neoscholatics" and his insistence that "It is of importance in the study of the human mind to turn to the mind itself rather than to the commentaries on Aristotle and St. Thomas" (p. 44).

The *second* part, on the Unconscious, closes with a brilliant discussion of the components of mind wherein Moore champions a tripartite division into functions, products, and dispositions. Pedagogically, his classificatory table of mental "elements" on p. 106 is one of the best of its kind.

Part II on the Emotional Life centers about a thorough demolition of the Lange-James theory. The author holds that "affections" are as much psychic as organic and that "the human being is a psychosomatic unit substance." "Faculties" are accepted as real entities on two grounds: (1) the necessary existence of mental "functions" to account for "products," and (2) the findings of factor analysis pointing to independent or distinguishable mental processes.

The *fourth* part is a rich treatment of representative psychopathological cases grouped under conventional categories. Here Moore's vast and solid learning, as evidenced by the extensive French, German, and Spanish footnotes, and his broad experience in "helping people in trouble" are prominent features.

The heart of this work appears in Part V which bears the same title as the book itself. It outlines a system of motivation with these emphases: "There are just as many impulses as there are human abilities"; "the criterion of worth in evaluating desires is not pleasure but accomplishment"; self-management requires a hierarchy of desires with one supreme end of life to which everything else must conform; the self-ideal is an ideo-motor concept; "religion is the natural sublimation of human desires, always possible and always effective, no matter how great the calamities that confront us"; man has many drives and not one. The twin chapters on the "parataxes" of depression and anxiety in this section are little works of art.

In Part VI the "Will" (with a capital W) is rehabilitated by arguing that if there is an admitted function of control, there must be a volitional faculty, which in turn is essentially a part of the "soul," or total organizing principle of man. Actually, this theory," while stressed, receives less space than extended analyses of experiments on motor response wherein Moore is thoroughly at home.

The *seventh* and concluding part is devoted to mental hygiene principles and a related philosophy of nature and society. There theology seems to win the upper hand, for the tone is basically that of a psychology of religion from the believer's standpoint.

Few books are harder to challenge than this one since it seems to be unusually free of factual errors. Certain judgments, of course, may be questioned, such as his belief that cyclothymia is transmitted in the genes, or the opinion that most of the "whining evaders of responsibility" he encountered in hospital wards had grown up without altruistic urges. One may also wonder whether there is not something spurious or at least weak in the (metaphysical ?) presuppositions behind his clever caricature of the literal superficiality of Wundtianism (or any other form of psychological positivism), viz: "There is a flow of consciousness, but what flows? Nothing. There is sensation and perception, but what perceives? Nothing. There is choice, but who chooses? Nobody. There is action, but nothing acts; motions and nothing moves" (p. 326). The field conception of the self-distribution of forces appears at the very end in a short but somewhat unintegrated discussion of the growth process, with the treatment implying that this position is all right as far as it goes, but that it doesn't go far enough. An astounding omission, too, in an earnest scientific testament of exceptional scope is the absence of any consideration of the cultural anthropologist's approach to personality structure.

Any one who finishes his perusal with the experience of "almost thou persuadest me" must acknowledge that this is a serious and important contribution to systematic psychology, of potential benefit to a wider audience than it will probably have.

GEORGE W. HARTMANN.

Teachers College, Columbia University.

POFFENBERGER, A. T. (ED.) *James McKeen Cattell: man of science. 1860—1944.* Lancaster, Pa.: Science Press, 1947. Vol. I, pp. viii+582; Vol. II, pp. viii+503.

Professor Poffenberger has rendered an important service to science and has paid tribute to a great scientist, teacher and apostle of academic freedom by bringing together in convenient form seventy of the more outstanding papers of James McKeen Cattell (1860–1944). The first volume contains twenty-nine of his scientific papers in which original researches are reported. In the second volume are forty-one papers, mostly addresses and semi-popular articles. Each volume has a short foreword by the editor followed by an appreciation of Cattell, in Volume I by R. S. Woodworth and in Volume II by F. L. Wells.

Cattell's complete bibliography, from which these seventy papers were selected, contains 167 titles which are listed in the back of Volume I. Of these 30 or more are reports of research work and constitute the bulk of Volume I. Those selected for the contents of Volume II are largely addresses, book reviews, notes and comments, most of which appeared in *Science* or *The Scientific Monthly*. Only two books are listed in the bibliography—*University Control*, published in 1913, and *Carnegie Pensions*, in 1919. Cattell believed that the best way to disseminate scientific information is not through books but by short papers in scientific and popular journals.

The editor does not state the basis on which the 41 papers for Volume II were selected. It appears, however, that they are representative of Cattell's broad interests in the advancement of science, its usefulness in daily life, and in his deep concern for academic freedom and democracy. More than half of the papers reprinted in Volume II are on psychology and its applications; about one quarter are on topics concerning the advancement of science in general; and one quarter on the organization and administration of colleges and universities in the United States.

In each volume the papers are arranged in chronological order and with no editorial change or comment. In view of the assertion by the editor that these papers have more than a mere historical value, one wonders why a topical arrangement was not employed. With the exception of two short papers (and possibly one other) the contents of Volume I could have been arranged under four main topics representative of the four major areas in which Cattell did his important research. These are:

1. *Studies of reaction times*, including measurements of the time taken up by the inertia of receptors, by sensation, perception, association, recognition, etc. At least ten of the papers in Volume I are reports of experiments in this general problem area.

2. *Studies in psychophysics*, particularly his famous work on the perception of small differences and measurement of errors of observation. The longest paper in Volume I—more than 100 pages—is his famous study, "On the perception of small differences," done in collaboration with Professor G. S. Fullerton of the University of Pennsylvania and published in 1892.

3. *The development of mental tests* and the measurement of individual differences. There are two papers on this topic in Volume I and three in Volume II.

4. *The development of the order of merit method* and its applications in his famous work on men of science. Ten papers comprising most of the last 200 pages of Volume I are on this general topic.

The papers in Volume II could have been classified as follows:

1. Those in which Cattell discusses technical problems in psychology, as for example, the papers on "The time it takes to think," "Reactions and perceptions," "The perception of light" and "Mental measurement."

2. Those in which he states his general viewpoint in psychology and his conceptions of psychological problems. For example, "The conceptions and methods of psychology."

3. Papers on the applications of psychology to practical problems, as for example, "Practical psychology," The usefulness of psychology," "Psychological methods to promote highway safety."

4. Papers on the organization of scientists and the advancement of science.

5. His extensive studies on the organization and control of American colleges and universities, including two or three papers on general education in relation to science and democracy.

6. And finally, two or three papers on wider social issues—democracy and international good will.

This outline will give the reader a general idea of the contents of these volumes. Psychologists will find in Volume I not only the original data of Cattell's early experiments but very stimulating discussions of problems and methods that are quite relevant to modern experimental work. This volume should and undoubtedly will appear on the reading lists for graduate students of psychology. It is a classic that will brighten with age. Cattell did as much as any contemporary psychologist to put psychology on a sound scientific basis and far more than any other man to establish it as a respected member of the growing family of sciences. He was the first man in the world to hold the title "Professor of Psychology" and the first psychologist to be admitted to the National Academy of Sciences.

Although Cattell never attempted to found a school of psychology or to formulate and promote a conceptual system, he none the less held rigidly to a behavioral point of view and regarded psychology as a growing and important science of behavior. "I have always held that psychology has to do with the conduct of an individual rather than with his consciousness. We are concerned with what a person does, rather than with what he thinks he thinks, feels he feels or imagines he imagines. In psychological experiments, as I once put it, 'It is usually no more necessary for the subject to be a psychologist than it is for the vivisected frog to be a physiologist'" (II, p. 377).

Although Cattell had little use personally for the introspective studies of consciousness and mental states, yet he maintained a healthy respect for all honest scientific efforts. He believed that psychology would grow as a science by the application of scientific methods, in the spirit of science, to its broad range of problems. In several places in his writings there appear statements like this: "The two great achievements of science have been the elaboration of the quantitative method on the one hand and of the genetic method on the other. . . . The physical sciences have been primarily quantitative and the biological sciences are primarily genetic . . . Psychology is from the start both quantitative and genetic" (II, pp. 203, 205). By the genetic method he meant the study of the previous conditions on which measurable activity depends. "Every mental state and every form of activity is the result of development from previous conditions. If explanation, as distinguished from description, is possible anywhere in science it is possible here" (II, p. 205).

Cattell never engaged in acrimonious controversies with other psychologists over points of view or scientific issues. His fighting talents, of which he had many, were used in other directions. They were aimed mainly at some university presidents, at boards of trustees, and particularly against the autocratic system of university control. The gallant fight that he led for academic freedom was prompted by his devotion to the three greatest things of life—science, democracy, and education; the greatest of which is science. These constitute "the triangular base on which rests the civilization that we have." They cannot be divorced. They must stand or fall together. They are centered in colleges and universities which are the sources of the greatest scientific achievements of our times. Universities must therefore, above all other institutions, be democratic in spirit and control. Cattell believed that one of the greatest stumbling blocks in the road of scientific and democratic progress, and hence in the path of human welfare and happiness, is the autocratic control of colleges and universities.

He tackled the problem in the democratic fighting spirit of Andrew Jackson. He loaded his gun with relevant facts charged with biting humor and fired them mainly at university presidents and boards of trustees. In his book on *University Control* he said, "In the academic jungle the president is my black beast . . . I once incited one of my children to call her doll Mr. President, on the esoteric ground that he would lie in any position in which he was placed. Of course, the president is by nature as truthful, honorable and kind as the rest of us, and is as likely to have more ability or enterprise, or both. But he really finds himself in an impossible situation" (II, p. 284). He goes on to propose certain reforms in university control which in his judgment would clear the way for the progress of science and the welfare of humanity.

The subtitle of the two volumes is "Man of Science." Cattell was all of this and much more. He was a great protagonist of science. Science to him was the most vital thing in life, amounting almost to a religion. Scattered through his writings, even in his most technical papers, are numerous references to the great importance of science for civilization and human happiness. It is the basis of democratic society; the discovery of truth that sets men free. It has freed them from toil and drudgery, from human slavery and child labor, from physical disease and pain, from torturing fears, and from "jealousies, resentments, and passions that lead to war." "Science has not only created our civilization: it has given to it the finest art and the truest faith" (II, p. 406).

Cattell was not a science-for-science's-sake man. For him science was not an indoor sport to be enjoyed by a chosen few. He regarded it as a powerful instrument of human welfare. His interest in applied psychology and his organization of The Psychological Corporation, to which he devoted the last years of his active life, bear eloquent testimony to this conviction. A characteristic statement of this belief is, " . . . it appears to me that the conditions are most healthful when science and its applications proceed hand in hand, as is now the case in engineering, electricity, chemistry, medicine, etc. If I did not believe that psychology affected conduct and could be applied in useful ways, I should regard my occupation as nearer to that of the professional chess-player or sword swallower than to that of the engineer or scientific physician" (II, p. 206).

There is an interesting parallel between the careers of Cattell and Benjamin Franklin. Both were interested in the advancement of science and particularly in its application to the practical problems of life. Both were journalists and devoted to the wide dissemination of knowledge. Each was the founder of a scientific society. Both were socially minded citizens: strong believers in the democratic process and bitter opponents of tyranny and regimentation, particularly in matters of scientific inquiry. There are, of course, important differences in the careers of these two great Americans, one of which may be changed as time goes on. The name of James McKeen Cattell may some day be as well known throughout the world as that of Benjamin Franklin.

MARK A. MAY.

Institute of Human Relations, Yale University.

GORER, GEOFFREY. *The American People. A study in national character.* New York: Norton, 1948. Pp. 246.

The scientific approach to the study of national character is beset by three major problems: what do we want to know, how are we going to get it, and how shall we interpret what we get? Cultural anthropologists interested in describing and explaining the general contours of small societies and their cultures have been able to overcome in some measure these problems faced by the scientist in the study of large, often heterogeneous, national groups. For, almost in one bold swoop, it has sometimes been possible for the culturologist to survey the social structure of a small tribe, describe its institutions and how they interlock with one another, make predictions as to how these institutions must inevitably affect the individuals involved as individuals and as a collective, and then to give case histories of a selected sample in order to show that the over-all description, explanation, and prediction square with the case material, i.e., with how the individuals actually think, feel, and act. These studies have for the most part approached the problem with psychoanalytic orientation so that not only are the assembled facts interpreted within this framework, but also the selection of facts demanded by the theory occupy the attention of the researcher, often at the expense of other ones equally important. Nevertheless, though these studies are really not as compact and all-explanatory and all-wise as they often seem when first one reads them, they have contributed greatly to our vision of the relations between the individual and his society, the flexibility and the basic stuff inherent in "human nature," as well as theoretical and methodological insights in the conception of "personality," "culture," "society," "institution," "social change," "normality," and "abnormality."

It is from this perspective that Geoffrey Gorer, like Margaret Mead[1] before him, has attempted to level his anthropological sights at the American people and to unearth their fundamental characteristics in one fell swoop. He was especially interested in making a contribution to the "mutual understanding and fruitful collaboration of the English and American peoples and governments, [because] such understanding and collaboration are our only safeguard against the inconceivable horrors of another war, or the horrors, only slightly less grim, of totalitarian domination" (p. 22). In this attempt, "only about two thirds of the inhabitants of the United States" are studied, thus excluding "the history, the traditions, and even to a large extent the population of the southern states and to a lesser degree Texas, rural New England, and California . . . " (p. 15). He further does not consider "the more obvious minorities, whether ethnic, religious, or social," within the confines of his vaguely delineated boundary-lines. Because of these considerations, one gets the feeling from the heart that not only are forty million Americans neglected in the study but that it is a study tailored-to-fit the "Anglo-Saxon" section whence the basic character of America is *supposedly* derived. This, of course, is a supposition which is taken altogether too much for granted by Gorer and can easily stand further investigation.

---

[1] MEAD, MARGARET. *And keep your powder dry.* New York: Morrow, 1942.

The study is largely impressionistic and based on Gorer's travels throughout the United States as well as his discussions with various social scientists and his "love and friendships and quarrels and misunderstandings and delicate negotiations and casual incidents which made up my life in the United States" during a period of seven years. Among a host of unrelated, though interesting, impressions and facts thus gained, Gorer presents a simple thesis which is supposed to "explain" what gives Americans their national character.

The essence of the thesis is not new; it is merely put in different words than others have done, especially Margaret Mead, an intellectual debt which he carefully acknowledges. In brief, we are told that the fundamental characteristic of Americans is their hatred of authority, and that this hatred stems from the fact that America, like no other country,[2] is a haven for immigrants, or at least has been until comparatively recently. First generation immigrants are depicted as having a strong hatred for their respective oppressors and having great hopes of adapting as best as possible to the new freedom of America, especially in the upbringing of the second generation as more perfect Americans than they are, in language, manners, appearance, aims, and values. "It is this break of continuity between the immigrants of the first generation and their children of the second generation which is to my mind of major importance in the development of the modern American character, which gave rise to what might be called by analogy with genetics, the American mutation. . . . The individual rejection of the European father as a model and a moral authority . . . was given significance and emphasis by its similarity to the rejection of England by which America became an independent nation" (pp. 26–27).

There can be no doubt that there is a great deal of truth in this theory of "Europe and the rejected father" as a basis for the national character of America, but when it is made to answer for almost every characteristic, it falls far short of the mark, as should be expected when one or two explanations are given for a set of infinitely complex and pluralistic social phenomena. When, for instance, we are told that it was only "the less assimilated and less assured groups" who loved Roosevelt like a father, and that the remainder had an "almost pathological hatred" for his authority, then it would seem that Dr. Gorer is overdramatizing the situation in order to fit it to his theory. It seems to this reviewer that an economist might more readily explain, and rightly so, the hatred for Roosevelt in terms of his position vis-à-vis the vested interests, and the love for him by the vast mass of the population precisely in terms of his authority in dealing fearlessly with these vested interests.

---

[2] Modern Palestine might be an exception to this statement, and to a lesser degree, Canada.

The crucial test of this theory is perhaps in the sphere of religion. While religious matters do not occupy the population nearly as much as in other countries, they are nevertheless, for the most part, God-fearing people endeavouring to do what is "right in the eyes of the Lord." "In God We Trust" is a common expression in the U. S., symbolizing that the American creed of "Life, Liberty, and the pursuit of Happiness" shall be guaranteed to all *under* God and *under* those whose position it is to administer this *maximum bonum.* Indeed, hatred of authority is manifested everywhere in the U. S., but it is in regard to interpersonal relations and not so much to all positions of authority in the land. It is perhaps because Gorer was not in a position to treat the religious question at all, as he points out in his Foreword, that he missed the significance of the "hatred of authority." It is for this reason that Gorer's Americans seem like a grumbling lot of authority-haters continually unable to distinguish between order and anarchy, between official leadership and self-appointed demagoguery, between duly constituted law and morals and the passing fancy of their individual tastes.

As the complement to the hatred of authority, i.e., the rejected father in psychoanalytic symbolism, is the inordinate love for the mother and things feminine. "The idiosyncratic feature of the American conscience [i.e., the superego] is that it is predominantly feminine," derived from the close ties between mother and child up to the age of six, and thenceforth reinforced by "school teachers who are overwhelmingly women." Business, politics, and the world of things (not persons) is exempt from this morality. For this reason of ambivalence, according to Gorer, "businessmen and their spokesmen" were violently opposed to the New Deal because "its social legislation was felt to be introducing into the domain of masculine privilege the meddling of female morality." Once again, one wonders what an economist or historian would say of such an explanation in their efforts to portray the Great Depression and its consequences. Accordingly, it would seem best for social scientists in the field of national character always to restrain themselves from giving a psychoanalytic explanation for large social issues, and only when no other explanation is forthcoming to feel free, though again with restraint, to resort to psychoanalytic or other psychological concepts. It is with such explanations in mind that Durkheim was prompted to write, and rightly so, many years ago: "Every time that a social phenomenon is directly explained by a psychological phenomenon, we may be sure that the explanation is false." Notwithstanding the corroboration from psychiatrists and other writers cited by Gorer, the theory of "mommism" may have its place in the consulting room or the clinic, but cannot, without being ridiculous, explain everything from the nursery and the school to the business office and the White House.

A similar criticism could be pointed out when Gorer claims that the "frequently expressed fears that America will be reduced to want, perhaps to actual starvation, if it lets its foods or resources or money outside the country," are due to the predetermined schedule-feeding of infants rather than on the basis of self-demand. Even if the "cause" for such occasional outbursts could be traced to the anxieties created by withholding the breast from the infant, it certainly does not jibe with the well-known fact of America's generosity to less-fortunate peoples, e.g., the aid given after the Japanese earthquake of 1923, the continuous sending of food parcels and money by all sections of the population to relatives overseas, and more recently the vast European Recovery Plan. "The great addiction of most American men to milk as a drink" as well as the "great erotic fetishistic value given to women's breasts" are also taken to be due to this early "anxiety." What is overlooked is that not only do they drink much milk but eat and drink many other things in greater quantities than elsewhere *because the standard of living is higher;* and further, with respect to the other claim, it has not yet been proven that European men are to any degree less erotically excited by women's breasts than American men.

On the question of "ideology" in the American character, Gorer does not have much to say, except to point out that it is "a minor component." While this assertion may be quite true, it seems likewise probable that it is true for the national character of most other peoples.[3] On the other hand, Northrup, Myrdal, and others have found a great deal of ideologism in the American character, so that to say that "ideology is a minor component of Americanism" is at best a controversial contention. Furthermore, some studies in the Psychological Barometer series would tend to show that Americans have pretty strong ideas (whether right or wrong is beside the point) regarding Americanism which guide their conduct, at least to some extent.[4]

Perhaps the most important contribution of the book is its novel analysis in bipolar terms of certain typical relationships: child to father, child to mother, parents to children, husband and wife, and so on. Though this type of analysis was first suggested by Gregory Bateson,[5] it is the first time as far as this reviewer is aware, that it has been as extensively applied in a study of American national character. In the

[3] Possible exceptions might be found among the French, Russians, and Israeli.

[4] Cf. for example, Henry C. Link, What does Americanism mean to the American people? *J. appl. Psychol.*, 1947, **31**, 425–430.

themal analysis of the various relationships he has chosen to study, Gorer gives us both an excellent structural and dynamic picture. Americans step out of print and somehow seem more real than they do in the rest of the book. It is a pity this analysis was not extended further.

In fine, notwithstanding the many criticisms that could be levelled at the book, Gorer deserves the thanks of all those who are interested in finding out what makes one people be what they are as compared to another people. Anyone who has thought and worked on this problem is well-ware of the difficulties involved in making generalizations which could easily be accepted by his co-workers. Like Newton's third law, to every generalization in national character there seems to be an equal and opposite generalization. This virgin field is wide-open for research and anyone who seriously attempts to close the wide gaps existing in our knowledge is making a big contribution. But a note of warning seems in place: analysis in depth (e.g., psychological interviewing, description, and analysis) can serve a very useful purpose only if it is then related to an analysis in breath (e.g., macroanalytic analysis of institutions and social structure as well as systematic polling and attitude studies) and if both are seen in the light of historical and contemporary social, political, and economic dynamics. Each approach separately tells only part of the story. Only an interlocked aggregate of the separate parts can give us the correct picture. This is a hard order but, in a world suffering greatly from international misunderstanding, the results would be worth the effort.

Percy Black.

Department of Sociology, University of Kentucky.

BOOKS AND MATERIALS RECEIVED

ALEXANDER, FRANZ, AND FRENCH, THOMAS M. *Studies in psychosomatic medicine.* New York: Ronald Press Co., 1948. Pp. xiii+568. $7.50.

BASLER, ROY P. *Sex, symbolism and psychology in literature.* New Brunswick: Rutgers Univ. Press, 1948.

BELL, J. E. K. *Projective techniques.* New York: Longmans, Green, 1948. Pp. xvi+533. $4.50.

BIERER, JOSHUA. *Therapeutic social clubs.* London: H. K. Lewis and Co., Ltd., 1948. Pp. 73.

BLAIR, GLENN, M. *Educational psychology, its development and present status.* Urbana: Univ. of Illinois Bulletin, 1948. Pp. 34.

BÜHLER, CHARLOTTE, BÜHLER, KARL, AND LEFEVER, D. W. *Development of the basic Rorschach score, with manual of directions.* 1948. (Lithographed by the authors.)

CHAMBERS, M. M. *Charters of philanthropies.* Boston: Merrymount Press, 1948. Pp. vii+247.

CHASE, STUART. *The proper study of mankind.* New York: Harper, 1948. Pp. xx+311. $3.00.

COMBS, A. W., *et al. Current trends in clinical psychology.* New York: New York Academy of Sciences, 1948. Pp. 867–928. $1.50.

CROW, L. D., AND CROW, A. *Educational psychology.* New York: American Book Co., 1948. Pp. xiv+599. $4.00.

DASGUPTA, JNANENDRA. *Behavior problems of school children.* Calcutta, India: Calcutta Univ. Press, 1948. Pp. xxv+110.

DENNIS, WAYNE. *Readings in the history of psychology.* New York: Appleton-Century-Crofts, 1948. Pp. 583. $4.75.

DUNBAR, FLANDERS. *Synopsis of psychosomatic diagnosis and treatment.* St. Louis: C. V. Mosby, 1948. Pp. 501. $6.50.

EGEL, PAULA F. *Technique of treatment for the cerebral palsy child.* St. Louis: C. V. Mosby, 1948. $3.50.

FARIS, R. E. L. *Social disorganization.* New York: Ronald Press, 1948. Pp. viii+481. $4.50.

FRANKLIN, JOHN F. *Annual report of the director of the clinic for child study for the year from July 1, 1946 to June 30, 1947.* Detroit: Probate Court, Wayne Co. Juvenile Division, 1947.

FROESCHELS, EMIL (Ed.) *Twentieth century speech and voice correction.* New York: Philosophical Library, 1948. Pp. ix+321. $6.00.

GATES, ARTHUR I., *et al. Educational psychology.* New York: Macmillan, 1948. Pp. xx+818. $4.25.

Goldstein, Kurt. *Language and language disturbances.* New York: Grune & Stratton, 1948. Pp. xii+374. $8.75.

Harmon, Francis L. *Understanding personality.* Milwaukee: Bruce Publishing Co., 1948. Pp. x+338. $3.50.

Johnson, Donald M. *Essentials of psychology.* New York: McGraw-Hill, 1948. Pp. xiii+491. $3.50.

Krech, David, and Crutchfield, Richard S. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948. Pp. xv+639. $4.50.

Law, Stanley G. *Therapy through interview.* New York: McGraw-Hill, 1948. Pp. xiii+313. $4.50.

Leuba, Clarence. *Ethics in sex conduct.* New York: Association Press, 1948. Pp. 164. $2.50.

Lewis, Don. *Quantitative methods in psychology.* Iowa City: Don Lewis, 1948.

Libs, Lester M. *Attitude prediction in labor relations.* A test of understanding.

Lief, Alfred. *The commonsense psychiatry of Dr. Adolf Meyer.* New York: McGraw-Hill, 1948. Pp. xvii+677. $6.50.

Munn, Norman L. *A laboratory manual in general experimental psychology.* (Rev. Ed.) Boston: Houghton Mifflin, 1948. Pp. xii+224. $2.50.

Patterson, C. H. *Using the Wechsler-Bellevue scales in counseling.* Minneapolis: In-Service Training Section, Personnel Division, VA Regional Office, 1948. Pp. 34.

Peterson, Harvey, A., *et al. Educational psychology.* New York: Macmillan, 1948. Pp. xiii+550. $4.00.

Pollack, Otto. *Social adjustment in old age.* New York: Social Science Research Council, 1948. Pp. x+199. $1.75.

Renard, Henri, *The philosophy of man.* Milwaukee: Bruce Publishing Co., 1948. Pp. x+238. $2.75.

Reusch, Jurgen, *Duodenal ulcer.* Berkeley: Univ. of California Press, 1948. Pp. vii+118. $4.00.

Rugh, Roberts. *Recent studies in the mechanisms of embryonic development.* New York: New York Academy of Sciences, 1948. Pp. 661+865.

Sachs, Hanns. *Masks of love and life.* Cambridge: Sci-Art Publishers, 1948. Pp. 303.

Sadler, William S. *A doctor talks to teen-agers.* St. Louis: C. V. Mosby, 1948. Pp. 379. $4.00.

Sartre, Jean-Paul. *The emotions. Outline of a theory.* New York: Philosophical Library, 1948. Pp. 97. $2.75.

SCHAEFER-SIMMERN, HENRY. *The unfolding of artistic activity.* Berkeley: Univ. of California Press, 1948. Pp. xiii+201. $5.00.

SIBLEY, ELBRIDGE. *The recruitment, selection, and training of social scientists.* Social Science Research Council Bulletin 58. New York: Social Science Research Council, 1948. Pp. xv+163.

STAGNER, ROSS. *Psychology of personality.* (2nd Ed.) New York: McGraw-Hill, 1948. Pp. xiii+485. $5.00.

SYMONDS, PERCIVAL M. *Picture-story test and manual.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1948. $5.25.

VAN LENNEP, D. J. *Four picture test.* Hague: Martinus Nijhoff N. V., 1948. Pp. 48. 65 Guilders.

VON HERTIG, HANS. *The criminal and his victim.* New Haven: Yale Univ. Press, 1948. $6.00.

WAGLEY, CHARLES. *Area research and training: A conference report on the study of world areas.* Social Science Research Council Pamphlet 6. New York: Social Science Research Council, 1948. Pp. v+58.

WHITE, ROBERT W. *The abnormal personality.* New York: Ronald Press, 1948. Pp. x+613. $5.00.

WIENER, NORBERT. *Cybernetics.* New York: John Wiley, 1948. Pp. 194. $3.00.

WOLFF, WERNER. *Diagrams of the unconscious.* New York: Grune & Stratton, 1948. Pp. xiv+423. $8.00.

WOODWORTH, R. S. *Contemporary schools of psychology.* (Rev. Ed.) New York: Ronald Press, 1948. Pp. ix+279. $3.00.

1948 Achievement testing program in independent schools and supplementary studies. New York: Educational Records Bureau, 1948.

Vol. 46, No. 2 March, 1949

Psychological Bulletin

THE PREVOST-FECHNER-BENHAM SUBJECTIVE COLORS

JOZEF COHEN AND DONALD A. GORDON
University of Illinois

Color vision, of all psychophysical phenomena, has perhaps the most elegant scientific formulation. The color manifold can be represented as a closed three dimensional space with reference to three arbitrary primaries. It thus fulfills the space requirements of solid analytical geometry and all transformation formulas referring to vectors in space can be made to apply.[1]

If the intensity function is eliminated by projecting the $x+y+z=1$ plane onto the xy plane, then the resulting configuration, called a chromaticity diagram, fulfills all requirements so that all theorems of projective geometry for closed spaces can be made to apply. We do not believe that there is any other psychological or biological function which may be so adequately described by mathematics so concise.

This remarkable fact has been noted by the mathematician, H. Weyl. With respect to the isomorphism between the projective plane and the sensation, he adds (**83**, p. 18),

> . . . . the projective plane and the color continuum are isomorphic with one another. Every theorem which is correct in one system . . . is transferred unchanged to the other . . . A science can never determine its subject matter except up to an isomorphic representation. The idea of isomorphism indicates the self-understood, insurmountable barrier of knowledge. It follows that toward the 'nature' of its objects science maintains complete indifference . . . For example, what distinguishes the colors from the points of the projective plane—one can only know in immediate intuition.

---

[1] This is actually the basis of modern color theory, but in recent years some workers have found it necessary to question the Euclidean nature of the space. This is stated implicitly by Troland (**79**) and explicitly by Silberstein (**69**, **70**). This latter series of papers presents a remarkable insight into what is involved in the mathematical considerations of the color domain. Silberstein asks whether the space is *Desarguesian*, and presents a striking technique whereby this concept may be put to empirical test.

The experimental fact that all colors may be reproduced by three real primaries involving negative coefficients, or three imaginary primaries involving only positive coefficients, leads to the inevitable hypothesis that there are three receptors in the retina, each of which is associated with a so-called "fundamental sensation curve." The enigma, the great enigma, of color vision is that *any* three points (not on a straight line) will serve as primaries, since lines connecting them will completely span the projective or chromaticity plane. Mathematically, any three points, corresponding to the primary colors, will produce a unique set of fundamental sensation curves, and are therefore as good as any other set of three points. If the object of science is prediction, then any set will predict color mixture data as well as any other set. However, if a choice of primaries can be made so that they predict something more than just the phenomena of color mixture, then they are desirable from the point of view of parsimony. People have, of course, done just this. Helmholtz selected his primaries on the basis of the fact that there are no changes in hue from 670 millimicrons on out to the long wave end of the spectrum, Hering on the basis of after images, and Ladd-Franklin on a phylogenetic basis. For a fine discussion on the nature and transformation of these primaries, see Judd (**48**).

There still remains an outside criterion for the selection of the primaries which has received little consideration from workers in color. We refer to the phenomenon of subjective colors, first discovered by Bénédict Prevost 122 years ago. We propose to present here a detailed history of the experimental work on these colors, followed by a brief discussion of the logic by which these colors may be made to determine the primary receptors and the corresponding projection and fundamental sensation curves.

The history of the phenomenon divides itself into several natural divisions, and the analysis will be in terms of these divisions.

I. PREVOST'S DISCOVERY AND THE FIRST PERIOD OF OBSERVATION

Bénédict Prevost (**62**), a French monk, discovered the subjective colors in 1826. He observed a ray of white light in a darkened chamber. He moved a white, rectangular cardboard perpendicularly to the ray, such that the ray was cut at successive intervals. He noted that the ray was white, at first, then a violet in the center, then a deep indigo, then a greenish-yellow followed by red. "One sees," he writes, "that the light of the ray . . . has been decomposed as if by a prism, into the seven principal colors ranging in almost the same order."

This is not to say that Prevost believed that he had actually decomposed the light as Newton did. On the contrary, he was aware that the effect was physiological in nature, and proposed a theory which has some degree of merit. Presumably, there are substances which combine with the nerves in the retina, giving rise to the sensations of color. These substances act at *different rates*. Since white light is composed of different wave-lengths, intermittent white light should produce the subjective colors seen. Since the concept of three receptors was unknown to Prevost, his theory compares most favorably with those of later more sophisticated investigators.

The first rediscovery of the subjective colors was made by G. T. Fechner (**33**) in 1838. It is Fechner who is usually credited with the original discovery. It was well known to Fechner that a disk whose composition is partially black and partially white will produce a gray when rotated, the luminosity of the gray being a function of the ratio of black to white. He noted, too, that the disk must rotate above the fusion frequency.

In order to produce various gradations of gray, Fechner prepared a disk which was divided into eighteen concentric circles, the diameter of each circle differing from the adjacent circle by a unit length. The innermost circle was completely black, the outermost circle was completely white. The remaining circles varied in the relative amounts of black and white as illustrated in Fig. 1a. The disk included in the Fechner paper shows only seven concentric circles. This was done so that the diagram might be clear, but Fechner never used such a disk. It is erroneously attributed to Fechner by many later investigators, for example, Colenbrander (**24**)

The circumference of the black figure (disregarding the corners) forms an Archimedes' spiral. While trying to spin the disk at a speed above the fusion frequency, Fechner spun it below the fusion frequency, and was amazed to see the many and varied colors which were produced. The colors, he notes, were not saturated, and yet were unmistakable. Other persons saw them in various degrees of distinctness. Some called the colors brilliant, some saw them only indistinctly, but all agreed that some color was evident.

Now Fechner begins to experiment with the possible determiners of subjective color. Speed of rotation and direction of rotation are factors. If white partial circles are used on a black background, as in Fig. 1b, colors are again produced although not in the same position as before.

If a disk of the type shown in Fig. 1c is rotated at high speed, a gray is produced as expected. At moderate rotation the disk color is yellow; at successively increasing speeds, it passes through yellowish-green,

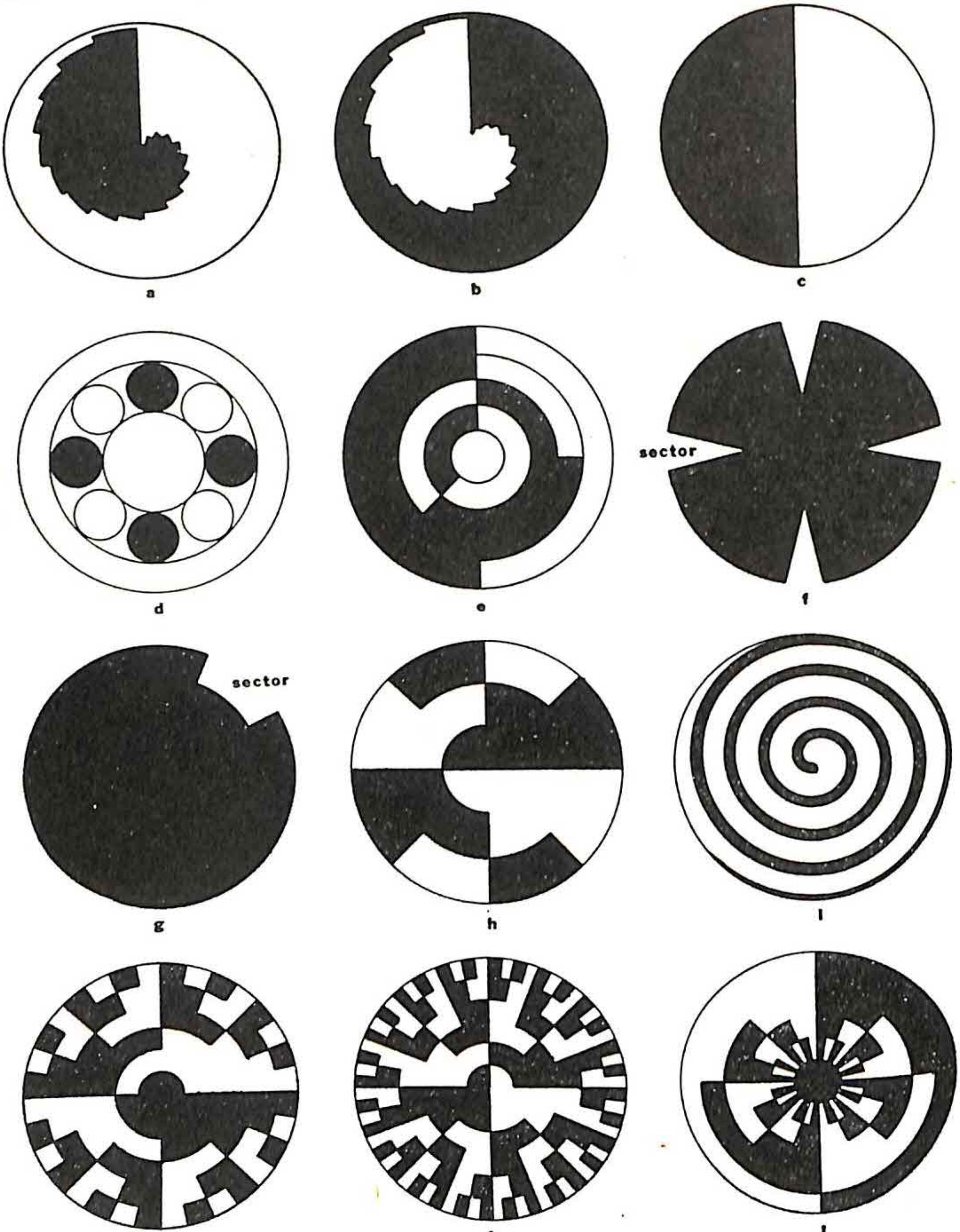


FIG. 1. DISKS FOR THE PRODUCTION OF THE SUBJECTIVE COLORS.
Figs. 1a, 1b, 1c, Fechner; Fig. 1d, Dove; Fig. 1e, John Smith; Figs. 1f, 1g, Rood; Figs. 1h, 1i, Helmholtz; Figs. 1j, 1k, 1l, Brücke.

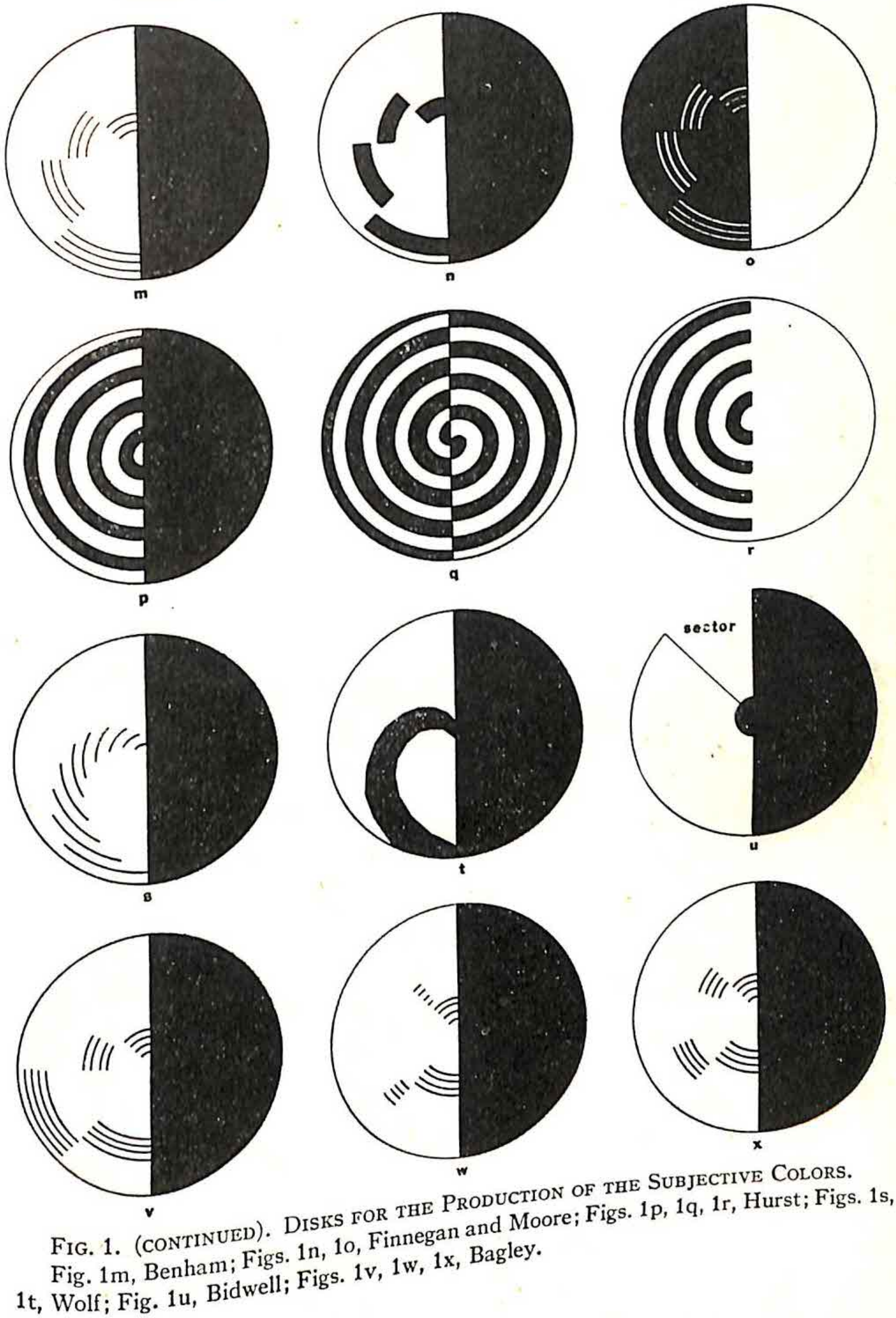


FIG. 1. (CONTINUED). DISKS FOR THE PRODUCTION OF THE SUBJECTIVE COLORS. Fig. 1m, Benham; Figs. 1n, 1o, Finnegan and Moore; Figs. 1p, 1q, 1r, Hurst; Figs. 1s, 1t, Wolf; Fig. 1u, Bidwell; Figs. 1v, 1w, 1x, Bagley.

green, and light blue. A "marbling" is interwoven through the colors appear on the disk. When the blue phase is reached, the whole disk is covered with a tapestry-like design of multi-colored cells. The nature of this design is suggested in Fig. 2. We have succeeded in reproducing this cellular design in the Illinois Laboratory in a clear and striking fashion. The drawing in Fig. 2 is not from Fechner, but from our own laboratory. The bases of the cellular formation are not clearly understood, and the existence of the formation has been noted, as will be seen, by many later investigators.

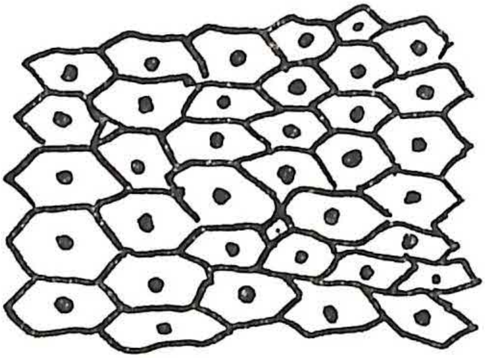


FIG. 2. CELLULAR DESIGN OF A HEXAGONAL NATURE ORIGINALLY REPORTED BY FECHNER.

The design accompanies the subjective colors. The drawing is from experiments performed at the Illinois Laboratory.

When the speed of the rotating disk is increased ever-so-slightly to the fusion frequency, the disk quickly fuses to gray.

Fechner advances a theory concerning the mechanism of the subjective colors. Imagine the disk to be rotating. When the observer fixates the disk, black replaces the white (which only an instant before replaced the black). The impression of the white does not disappear immediately, and, as a matter of fact, lasts for some time as is demonstrated by, as Fechner says, "the phenomenon of after sensations." The perception of white is dependent upon many receptors in the eye, probably three. But the rate of fall (Fechner does not mention the rise, but should) of each of these receptors is different. And it is this differential fall (and rise) which is the basis of the sensation of color. A moment later the eye is stimulated by white light again, the white is replaced by black, and the entire cycle repeats itself.

Dove (**26**, **27**) in 1848 noted the relation between Fechner's colors and those produced as a simple after sensation. He describes disks similar to those of Fechner, with the difference that white is replaced by another color. When the disks are rotated he describes the resulting perception. The paper is of theoretical value since it demonstrates the relation between after sensations and subjective color—that all colors are really subjective—but makes no really significant contribution. Dove's disk is reproduced, after his description, in Fig. 1d.

In 1859 John Smith (**73, 74**) made *the second rediscovery*. He believed that the subjective colors were a direct result of the physical characteristics of light and that the new theory of light which he proposed would replace that of Newton. We reproduce verbatim a portion of Smith's abstract.

> The author produced the same results by cutting out spaces in a white card, and causing it to revolve on a black surface. He produced also similar phenomena by causing these figures to revolve when held perpendicularly, and to take the appearance of coloured solids. He also caused these colours to be reflected on a white surface from the revolving disk.
>
> Remarkable as these experiments are, they are not more remarkable than the results they lead to.
>
> They prove the homogeneity of the ether.
>
> They prove the undulatory hypothesis, but oppose the undulatory theory.
>
> They show the necessity of introducing a negative element into the theory of colour, or that colour is the effect of two coordinate sensations—a positive and negative.
>
> They enable us to dispense with the different refrangibilities of the rays of light, taught by Newton.
>
> They remove the necessity for the supposition of different waves or of a disposition in matter to produce waves of different lengths.
>
> They help to explain the phenomena of what is called the polarization of light.
>
> They give a new explanation of prismatic refraction, and explain in a plain and simple manner many very interesting natural phenomena.

Smith indicates that there are five rings on his disk, but describes only four. Fig. 1e illustrates J. Smith's disk, reconstructed from his description. One may conclude that Smith had perhaps misinterpreted the theoretical aspects of his rediscovery of the subjective colors.

Rood (**63**) in 1860, realized that the theory of light proposed by John Smith was invalid, since the color appearances were subjective. He confirmed the existence of the colors, but disagreed as to their origin.

Rood first observed the subjective colors while attempting to measure the time of small explosions of gunpowder by use of a rotating disk. For this purpose he used a black disk with four radial slits, reproduced in Fig. 1f. A certain speed caused the lower part of the flame to appear green, while at slower speeds the entire flame appeared to be a purplish-red. At very slow speeds of rotation, the appearance of the flame alternated between white and violet.

A more elaborate disk of the type illustrated in Fig. 1g was now constructed. It was nine inches in diameter, while the slit was one half inch at the circumference. The disk was rotated against a white cloud. At a rotational speed of ten revolutions per second, the cloud appeared to be a deep red.

The disk was viewed through a colored filter whose transmission properties were such that it transmitted the red, orange, yellow, some green, but excluded the blue and violet. If radiations coming from the disk were red and only red, then the filter should transmit these rays and the resulting sensation should be red. The sensation, under these conditions, is not red but a greenish-blue, indicating the subjective nature of the phenomenon. The subjective nature is further shown in that if a yellow glass is employed, the disk appears violet; if a green glass be used, the disk appears green, neutral, or faint red.

Brewster (16) in 1861 made *the third rediscovery* of the subjective colors before becoming familiar with the work of Prevost and John Smith. He observed the colors while watching a rotating series of slits pulse a white light source. The colors, beginning at the center of his observational field, were as follows: white or bluish, darker blue, white, a well defined dark ring, white, greenish yellow, reddish. He suggests that similar phenomena may be seen in a flickering gas flame.

Using an apparatus consisting of a rotating disk in front of an illuminated ground glass, he was able to reproduce the cellular, hexagonal pattern reported by Fechner. He reports, too, an additional quadrangular ". . . pattern, which is too evanescent to permit it to be drawn, [which] consists of a series of dark quadrangular spaces separated by triple or multiple lines of light. It has no relation whatever to the hexagonal pattern, and has never been seen unless the eye has been strongly impressed by the successive impulses of the luminous disk." He suggests that the hexagonal forms may represent cells in the choroid coat; he is uncertain concerning the origin of the quadrangular forms.

Rood (64) published a short note in 1861 in which he reviewed the meager literature on subjective color then available. He suggested that the hexagonal pattern reported by Brewster might be related to similar observations reported by Purkinje and Vierordt (see 43, p. 257).

II. HELMHOLTZ TO STEWART—THE SCIENTIFIC FORMULATION

The genius of Helmholtz began the second well defined historical period. Here we find the first formal applications of scientific principles. Helmholtz (43) in 1860 (translated in 1924) pointed out that if an observer fixates a Maxwell disk with black and white sectors, he notices that the white is red on the advancing edge and bluish on the other edge. In reduced illumination, the red becomes reddish yellow, and the blue becomes violet. In intense illumination, the red becomes pinker, and the blue becomes greenish. "With slower rotation," he writes, "the bluish hue at first is spread over a wider part of the white than the reddish. But when the speed is faster, red spreads out all over the white as pink-red, and green-blue extends over into the black sectors; and on the whole, violet appears to predominate on the disk." These three phases can be demonstrated simultaneously by use of a disk reproduced in Fig. 1h. The colors can be seen somewhat more easily by use of the disk in Fig. 1i.

Practice is necessary to see the colors in their maximum saturation. Apparently, the eye must be fatigued by the flicker to arrive at the optimum conditions. He reports the hexagonal cells. If a partially colored and partially black disk is rotated, the observed colors tend to fade and be non-homogeneous. He believes that the "fluttering heart" illusion is related to the subjective colors.

Helmholtz' theory is that of Fechner—that the three receptors have a differential rise and fall. The blue, says Helmholtz, lags behind the red.

Brücke (19) utilized disks illustrated in Figs. 1j, 1k, and 1l. He proposed to show that Talbot's Law, as a result of Fechner's experiments in subjective color, was not true under all conditions. Following a theoretical discussion of the relation between the objective stimulus and the resulting sensation, he spun his disks in order that he determine which speed produced the brightest appearance. The brightest color, obtained when violet just about goes over to blue, was observed at an average rate of 17.6 white pulses per second.

He concluded that the perception of the subjective colors was closely allied to what modern investigators call adaptation or fatigue, *i.e.*, the perception brought about by the steady stimulation of the retina. Actually, by definition, the subjective colors are produced only by intermittent light. For example, after images, although really as subjective as any other color experience, are not subjective in this sense. The entire discussion is confused by the interchange of what we would call the adaptation or fatigue of the receptor and the fall of the receptor.

Aubert (4) included a discussion of the subjective colors in his famous book on physiological optics in 1865. He indicates that he sees the colors on Brücke's disk at the greatest saturation after a fixation of about one half minute. The speed of the disk is a determiner of the colors seen. If the illumination of the disk is decreased, then the speed of rotation of the disk must be increased in order that the appearance remain constant.

With reference to the Brücke disk (Fig. 1j), he says that if the frequency of the white pulses is more than 56 per second, then the ring appears homogeneously gray. If the pulses are between 42 and 48 per second, and an outside part of the ring is fixated, then the part in direct vision will be gray while the part in indirect vision will appear to be a Berlin blue. As the speed decreases the extension of gray in direct vision decreases. The ring does not appear entirely uniform, being interposed with dark and light radial stripes, and blue and gold become apparent. As the speed is decreased the blue and gold become more intense and the square (hexagonal?) cell structures are seen. As the speed of the disk continues to decrease the extent of the blue field decreases until it disappears at 10 pulses per second. Various other color phenomena are described.

Exner (**32**) in 1870 also observed that the subjective colors appear to differ when viewed in the center or the periphery of the eye. His introspective descriptions are given in great detail. The colors, he points out, are seen in greater saturations when the eye is fatigued. But the fatigue of one eye does not facilitate the perception of subjective colors in the other eye, and it would follow, therefore, that this aspect of the subjective colors is not a central effect.

Exner reported the presence of quadratic or six sided pointed elements, and expressed his belief that an explanation of them would be of fundamental importance to physiological optics.

F. J. Smith (**72**) made *the fourth rediscovery* in 1881. He theorized that light might be decomposed by interruptions proportional to the wave-length of a particular ray of the many, which together are the composite ray. He used a large bicycle wheel between the sun as a light source and the observer, and produced the subjective colors. Performing simple experiments he believed that he had derived the wavelengths of the spectral colors.

The fifth rediscovery was made by Hannay (**42**) in 1881. While working on dry photographic plates in a room illuminated by a ruby light, he noticed that movements of his fingers produced a greenish blue, with blue predominating. He independently derived the Fechnerian theory of the rise and fall of the receptors, and suggested that this phenomenon might be used as a basis for determining the fundamental color sensations.

N. Smith (**75**) stated in 1882 that Hannay's explanation (the Fechnerian theory) was suitable for colors seen in the photographic dark-

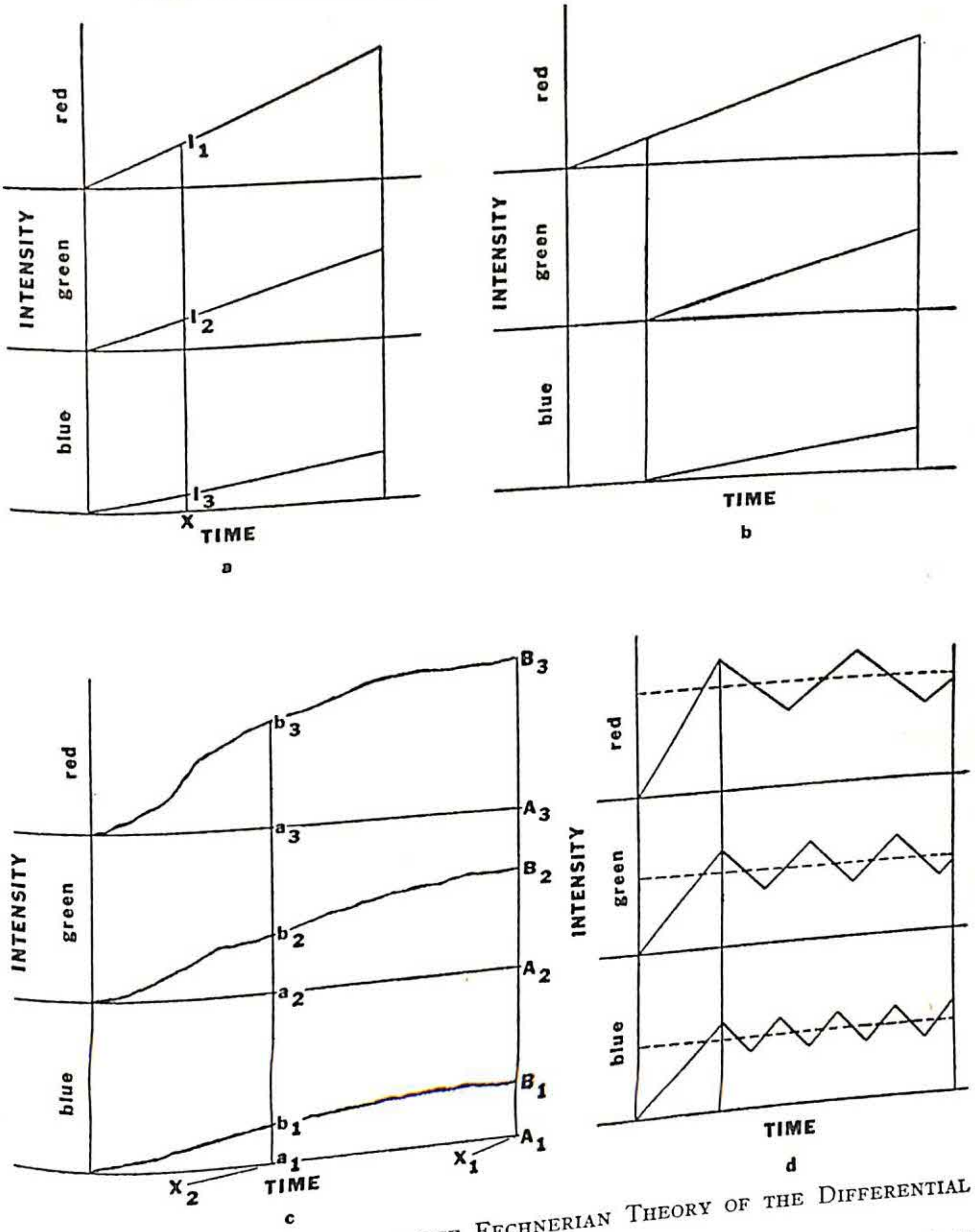


FIG. 3. Stewart's ANALYSIS OF THE FECHNERIAN THEORY OF THE DIFFERENTIAL RISE AND FALL OF THE THREE RECEPTORS.

Fig. 3a illustrates the case where the rise is linear, the slopes are different, and all three receptors begin the rise at the same time. Fig. 3b illustrates the case where the rise is linear, the slopes are different, and the three receptors begin their rises at different times. Fig. 3c illustrates the case where the rise is non-linear, the slopes are different, and all three receptors begin the rise at the same time. This last case is shown to be the most adequate. Fig. 3d indicates the responses of the three receptors under pulsing white light—pulses which are of such short duration as to be rarely perceived as white.

room, but was unsatisfactory for colors seen on a rotating disk. He offered no alternative explanation.

Stewart (76) made *the sixth rediscovery* in 1887. Investigating Talbot's Law, he noticed that observation of a rotating mirror gave rise to the subjective colors. He described the results of careful observations under varying lights of different intensities.

Stewart independently derived the Fechnerian theory of the differential rise and fall of the color receptors. Further, he determined the mathematical nature of the different response curves from his experimental results, and derived the conditions under which they must operate. Because of their great importance, his proofs are summarized here. Stewart indicates that there are three possibilities concerning the nature of the curves:

> *Case 1* is illustrated in Fig. 3a, where the rise of each of the three receptors is linear, all three begin at the same instant, but the rate of rise for each of the three is different. The curves are shown on three separate horizontal axes, but actually they are superimposed one on the other. The color seen at any given instant, say at time X, is the ratio $I_1:I_2:I_3$. When I_1, I_2, I_3 each reach their maximum, then white is seen. For this case, although different colors can be seen depending on the ratio $I_1:I_2:I_3$, the color will not change as the receptors reach their peak of excitation. Experimentally, the subjective colors do change in time, and case 1 is therefore impossible.
>
> *Case 2* is illustrated in Fig. 3b, where the rise of each of the three receptors is linear, but all three do not begin at the same instant, and again the rate of rise for each of the three receptors is different. It is obvious that subjective colors may arise under these conditions, since one receptor may be responding while the other receptors are not responding. It would be a more powerful proof if the assumption of linearity could be eliminated, bringing us to the next case.
>
> *Case 3*, where the rise of each of the three receptors is non-linear and, it is assumed that all three begin at the same instant, is illustrated in Fig. 3c. The color (which in this instance is white) seen at time X_1 will be given by the ratio $A_1B_1:A_2B_2:A_3B_3$. The color (which is not white) will be different when the ratio is $a_1b_1:a_2b_2:a_3b_3$ at time X_2, since these ratios differ from the former ratios. This explains the mechanism of the subjective colors.
>
> Fig. 3d indicates the responses of three receptors under pulsing white light—pulses which are of such short duration so that the stimulus is almost never perceived as white. The assumption is made that the pulsing light sets up different phase relationships among the receptors, since the natural periods of each of the receptors is probably different. Observations indicate that as the intensity of the pulsing white light is increased, the subjective colors tend toward violet. The explanation is found with reference to Fig. 3d, where it is assumed that the change in rate of rise is most appreciable in those curves which, at a given speed, are less steep.

Stewart reports many experiments in which colors (some monochromatic) other than white were pulsed. He reports many subjective colors, which indicates that more than one receptor is being stimulated.

Charpentier (**21**) rediscovered the subjective colors for *the seventh time* in 1891. He observed the phenomenon by looking at light which was intermittently pulsed through holes. In a second paper (**22**) he discussed the theoretical implications of his "discovery." He reports that intensity of light is not a factor. He states that the subjective colors tend to refute the Helmholtz trichromatic theory of color vision, because, all colors are reproduced, not just three. The alternative theory which he proposes is not reasonable. In a third paper Charpentier (**23**) describes an additional technique for producing the colors. The paper is not of great importance.

Sanford (**65**) was apparently the first individual to include the subjective colors in classroom teaching. He discussed an experiment on the colors in his course in physiological psychology in 1893.

III. REDISCOVERIES IN 1894–95. THE REVERSION TO OBSERVATION

In the years 1894 and 1895 there began in England a period of considerable interest in what was called the artificial spectrum top. The discussions appeared in the British journal *Nature*, and were keynoted by a description of the so-called Benham top (**84**), illustrated in Fig. 1m. This top, sold extensively as a toy, presents colors in an extremely striking fashion—and is even more interesting since the position of the colors is reversed when the direction of rotation is reversed. The Benham disk constituted *the eighth rediscovery.*

Liveing (**52**) presented an address to the Philosophical Society of Cambridge in 1894 in which he presented a variety of disks similar to that of Benham and presented a theory very similar to that of Fechner.

Benham (**9**) joined the discussion by saying that Liveing's theory was wrong and his own correct, since the colors could be seen under the illumination of a sodium lamp. It was not made clear as to why the Fechnerian theory would not account for the colors under these conditions.

Liveing (**53**) replied that he had observed the top under the sodium light, and had observed colors, although the colors were not the same as those observed under white light. He too fails to state that the Fechnerian theory will account for the colors, but merely repeats that the colors are of a subjective nature.

Abney (**2**) rotated the Benham top in white and near-monochromatic lights. The following colors were observed, where No. 1, No. 2, No. 3, and No. 4 are the triple lines of the top (see Fig. 1m) in order from the center out:

With white light: No. 1 crimson, No. 2 olive green, No. 3 gray (slightly violet), No. 4 dark violet; *with red (C light)*: No. 1 red, No. 2 lighter red, No. 3 very light olive green, No. 4 darker olive green; *with green (Magnesium b)*: No. 1 bluish green, No. 2 lighter bluish green, No. 3 lighter bluish green, No. 4 ruddy black; *with blue (near the blue Lithium)*: No. 1 grass green, No. 2 lighter grass green, No. 3 lighter grass green, No. 4 ruddy black; *with violet (all the violet of the spectrum)*: No. 1 light violet, No. 2 light violet, No. 3 light violet, No. 4 darker violet with a suspicion of red.

Abney states, "When a red a little below the red lithium line was employed, all the groups appeared dark red, and, as in the three sensation theory this part and the violet are simple sensations, the results obtained in these last were to be expected."

Continuing with mixtures of two simple colors, the following results were obtained:

With a mixture of red and green to make white: No. 1 indigo blue, No. 2 reddish orange, No. 3 reddish orange, No. 4 darker orange; *with a mixture of yellow and blue to make white:* No. 1 sky blue, No. 2 sage green, No. 3 sage green, No. 4 bluish black (perhaps black).

These results are extremely important in the theoretical application of these colors to the isolation of the primary receptors.

Finnegan and Moore (**34**) constructed two disks illustrated in Figs. 1n and 1o. They found that the width of the lines was highly important. If the lines are too broad, then color appears only on the edge. Further, the sequence of black and white with respect to the lines and the rate of rotation are determiners of the colors seen.

Edridge-Green (**29**) pointed out that the subjective colors of the Benham top were previously discussed by Helmholtz and Fechner. He reports the hexagonal cell structure when the Benham top is rotated.

Benham (**10**) mentioned in a note that his theory (which is improbable) would adequately account for the change in colors under change in rate of rotation mentioned by Finnegan and Moore. The statement is not clear.

Turner (**80**) suggested an apparatus whereby a glass spectrum top is inserted in the slide holder of a projector, rotated and projected, so that demonstrations might be made to large audiences. He also proposed the use of colored filters for additional effects.

Newton and Co. (**55**) called the attention of the scientific public to the fact that the Benham disk was copyrighted and that the use of the disk for projection was prohibited.

Hurst (**45**) then suggested three new disks which were not copyrighted. They appear in Figs. 1p, 1q, and 1r.

A note appearing in the British journal, *Engineering* (**86**), made mention of the Benham top. A theory of one MacFarlane Gray is discussed. According to Mr. Gray, the subjective colors are caused by the color aberration of the lens of the eye. He proposes an "eye-demon" which compensates the lens, under normal conditions, for every change in seen color. When compensation cannot be made quickly, subjective colors are seen.

These notes were followed by still another in the *Scientific American* (85). The Benham top is again presented. MacFarlane Gray's theory is discussed with the statement, "It does not, perhaps, present the degree of clearness and precision to which we are accustomed in the study of optics. The field remains open to investigation for varying the experiments and completing the first exposed." Two new disks, designed by Charles E. Wolf, were presented and are illustrated in Figs. 1s and 1t. The disk in Fig. 1s is the obvious extension of the Benham disk, since the lines will produce finer gradations of color. Fig. 1t is the same as Fig. 1s, with the difference that the lines have been connected so as to produce continuous variations in color.

James McKeen Cattell (20) concluded the 1895 series of notes by pointing out that the spectrum tops were introduced to modern science by Fechner, Brücke, and Helmholtz, and hence were not patentable.

Henry (44) in 1896 attempted to explain the colors of the Benham top by the differential sensitivity of the periphery and fovea of the retina. He even went so far as to determine the different sensitivities of the foveas of several individuals. This is not the mechanism of the colors, as will soon be seen from the experiments of Doniselli and Piéron.

In 1896, Schwartze (66), in an appendix to his book on molecular physics, attempted to show that "the Newtonian hypothesis . . . is based on the absurd assumption of the multicoloration of the unicolored light," by use of Fechner's disk. This he did unsuccessfully.

Bidwell (11) in 1896 mentions several new methods of producing subjective colors. These are somewhat similar to those of Charpentier, to whom Bidwell makes reference. Bidwell believes that the colors of the Benham top may be accounted for by his theory that color fibres adjacent to those of the stimulated area of the retina continue to respond after a stimulus has ceased, by a process of "sympathetic reaction."

In a second paper Bidwell (12) in 1897 stated that the greenish blue subjective color observed in his previous experiments was due to the after image of the subjective red color. This is rather remarkable for, as Bidwell himself states, the greenish blue may be much more intense than the red. He made the additional observation that the illumination is a determiner of the color seen. He found, further, that an observer may not be able to distinguish between a real and a subjective red. In the following year, Bidwell (**13**) reviewed his experiments and published a picture of his disk which appears in **Fig. 1u.**

A note then appeared in the French journal *La Nature* (**41**). The colors were demonstrated by means of a half black, half white disk, with a triangular cut out (similar to Bidwell, see Fig. 1u). When rotated in front of a red field, one sees blue-green instead of red. The theory is advanced that the colors are due to retinal fatigue.

IV. BAGLEY AND DONISELLI. THE PSYCHOPHYSICAL APPROACH

In 1902, Bagley (**5**) began a new period in the history of subjective studies with the first, and one of the best, systematic investigations of the phenomenon. Included with this study is the first extensive historical survey.

Miss Bagley utilized ninety-seven modifications of the Benham disk, three of which are illustrated in Figs. 1v, 1w, and 1x. Systematically, Bagley studied the effects of the length of lines on the color seen, the effect of the width of the line on the color seen, the effect of the distribution of the groups of lines on the disk, the effect of background color, the effect of the amount and distribution of background color.

Spinning an ordinary Benham disk, she reports that black followed immediately by white shows the lines to be red. When black is both followed and preceded by white, the lines appear to be green. Variations of the amounts of white vary the color of the green. Black preceded by white shows the lines to be blue. When lines which appear to be red green, and blue, under normal rotation are lengthened, the colors appear to be less intense but not less saturated. As a general rule, the width of the lines has no effect on the hue seen, but it may effect the other color variables.

When the black sector is varied, experiments show that the colors seen are not dependent merely upon position, but upon actual duration of the stimulation. With an increase of black, the reds are emphasized. When white is increased, the blues are emphasized. In bright daylight, the blue is emphasized. In less intense artificial light, the reds are emphasized.

Reversal of rotation of the Benham disk does not give an exact reversal of the colors seen. The hues are reversed, but the colors appear in different saturations and intensities.

When a gray screen is placed about a particular ring of one of the disks, the effect is to decrease the color seen of the ring. When the white

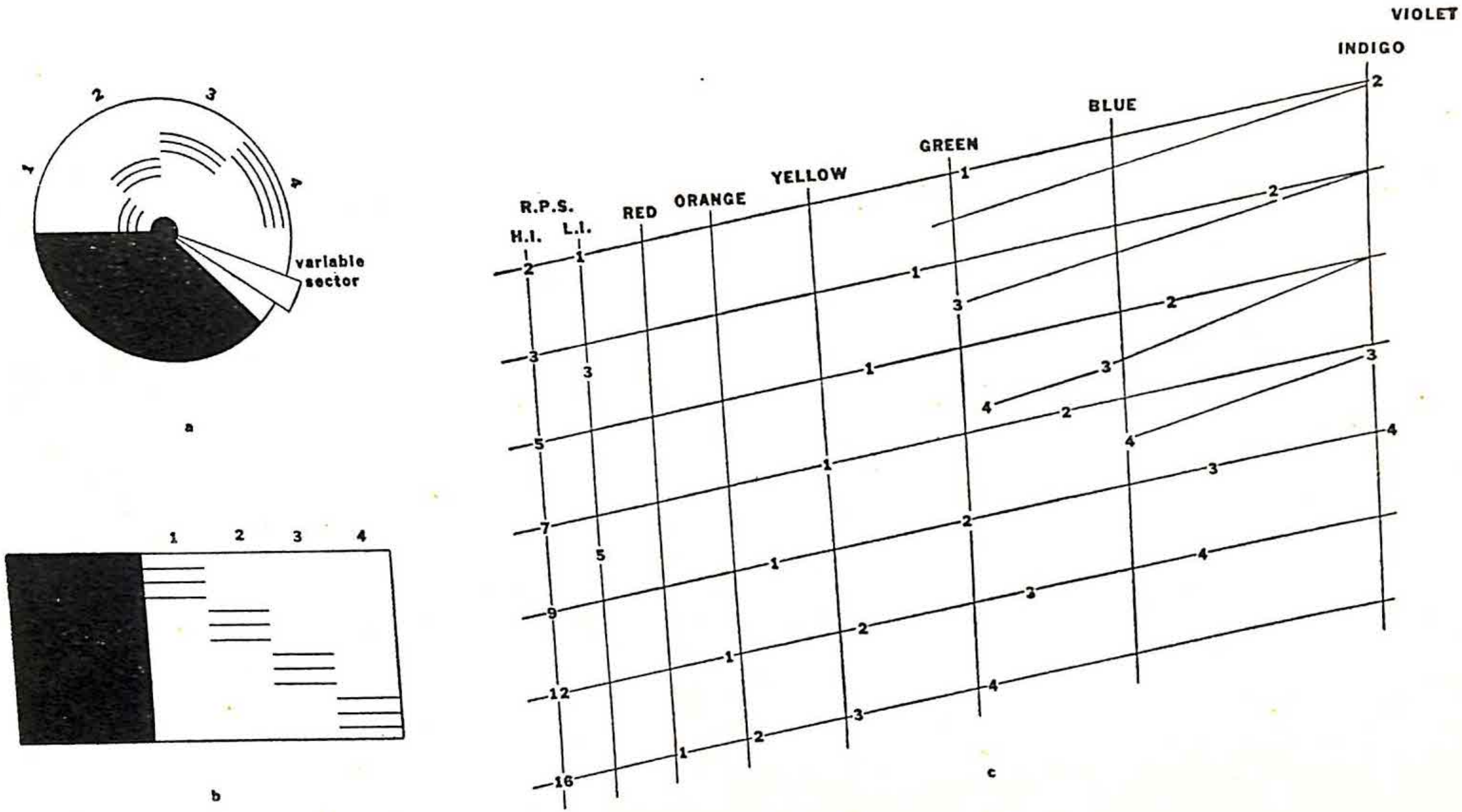


FIG. 4. DONISELLI'S ANALYSIS OF THE SUBJECTIVE COLORS.

FIG. 4a illustrates a Benham top which has a variable black sector. Fig. 4b is the analogue of Fig. 4a—a design to be used on a kymograph drum. Fig. 4c is a redrawing of Doniselli's results. The numbers refer to the arcs in Figs. 4a and 4b. The colors seen on the arcs are shown to be a function of the speed of rotation and the level of illumination. R.P.S. = Rotations Per Second. H.I. = High Intensity. L.I. = Low Intensity.

of the Benham disk is replaced by a physical color, very interesting results are obtained. Lines which would appear red if on a white background, appear green if on a red background; lines which would appear blue if on a white background, appear to be reddish brown on a dull orange background. These experiments illustrate the contrast effects possible with subjective colors.

Bagley notes some of the general conditions of observation. Short periods of keen attention aid in the perception of colors. Visual fatigue tends to deaden the colors. Practice, contrary to Helmholtz, is not necessary to see the colors. There is an optimum speed of rotation at which the colors are seen most clearly.

Miss Bagley relates her results to the Ebbinghaus theory of vision. Her theoretical discussion is not clear and will not be reviewed here.

Doniselli (25) in 1907 published the second of the important experimental papers. He constructed a Benham disk having a 225° white sector, a 135° black sector, and a variable white section which allowed the black to be reduced to $\frac{1}{5}$ of the white section. The disk and its analogue, a design to appear on a kymograph drum, are illustrated in Figs. 4a and 4b. Using this variable disk he systematically investigated the effect of the black/white ratio, the speed of rotation (2–17 revolutions per second), and illumination.

Doniselli presents his results in a systematic form. They are shown redrawn in Fig. 4c. Apparently, the color seen is, in most cases, a linear function of the speed of rotation. At greatest speeds, the colors in the rings correspond to the natural order of the spectrum. At lower speeds there are inversions.

In his theorizing concerning the cause of the phenomenon, Doniselli writes as follow:

> a. Every sudden stimulation by intense white light of brief duration, gives rise to a primary impression which is predominantly red-yellow.
>
> b. The primary red-yellow phase is followed by a secondary blue phase. The speed of passage from the primary to the secondary phase is directly proportional to the intensity of the original white stimulus.
>
> c. When the intensity is high enough, the primary phase is too rapid to be seen, and the colors seen appear to be reversed relative to those seen at medium intensities.

Doniselli presents a mechanism for the subjective colors which is a direct analogue to a vibrating string varying in tension, which produces tones. Doniselli has done what so many other investigators have done—he has confused the subjective colors with colored after images. Finally Doniselli eliminated the hypothesis of Henry that the subjective colors were caused by the differential distribution of the cones in the retina, by demonstrating the colors on a sheet of paper rolled about a cylinder (Fig. 4b).

Sexton (**68**) discussed the Benham top in 1907. He indicated that the color was affected by the position of the sector lines, the length and width of these lines, the direction of rotation and the illumination. He independently derived the Fechnerian theory of the differential rise and fall of the receptors. He observed that the fact that colors are seen in sodium light in no way disproves the Fechnerian theory—since the sodium light affects the three receptors.

Percival (**57**) in 1909 became interested in the problem, but simply repeated Stewart's experiments and presented Stewart's diagrams and theories.

Baumann (**6**) published the first of three papers on the subject in 1912. He described a simple apparatus for rotating the disks, and described the colors of the Benham disk which he erroneously attributes to Schwartze. He presents three new disks. In a second paper (**7**), he states that the smaller the image of the disk upon the retina, the more vivid are the subjective colors. In his last paper (**8**), published in 1918, he presented five new disks which are minor variations on Wolf's tops published in the *Scientific American* (see Figs. 1s and 1t), to which he makes no reference. Factors which influence the colors seen are the nature of the eye, the size of the sectors, the speed of rotation, and the illumination. In direct sunlight he sees no colors at all. He proposes a theory based on the concept that nerves, although conducting the colors seen, cannot conduct more than a certain load. The theory would appear to be improbable

Edridge-Green (**30**) in 1916 reported that the subjective colors could easily be seen on a black and white design rotating on a kymograph drum. The colors, he says, are the result of after images. It is difficult to understand how, as he says "the after image of black is green."

V. Interest in Non-White Stimuli

In 1917 and 1918 Ives wrote two papers which, together, constitute *the ninth rediscovery* of the subjective colors. Actually, they constitute one of the really great experimental contributions to the field of color, but Ives was not only unaware of their importance but was completely amiss concerning the theoretical implications of his findings. This is all the more interesting, since Ives was a man of monumental proportions in the field of physiological optics.

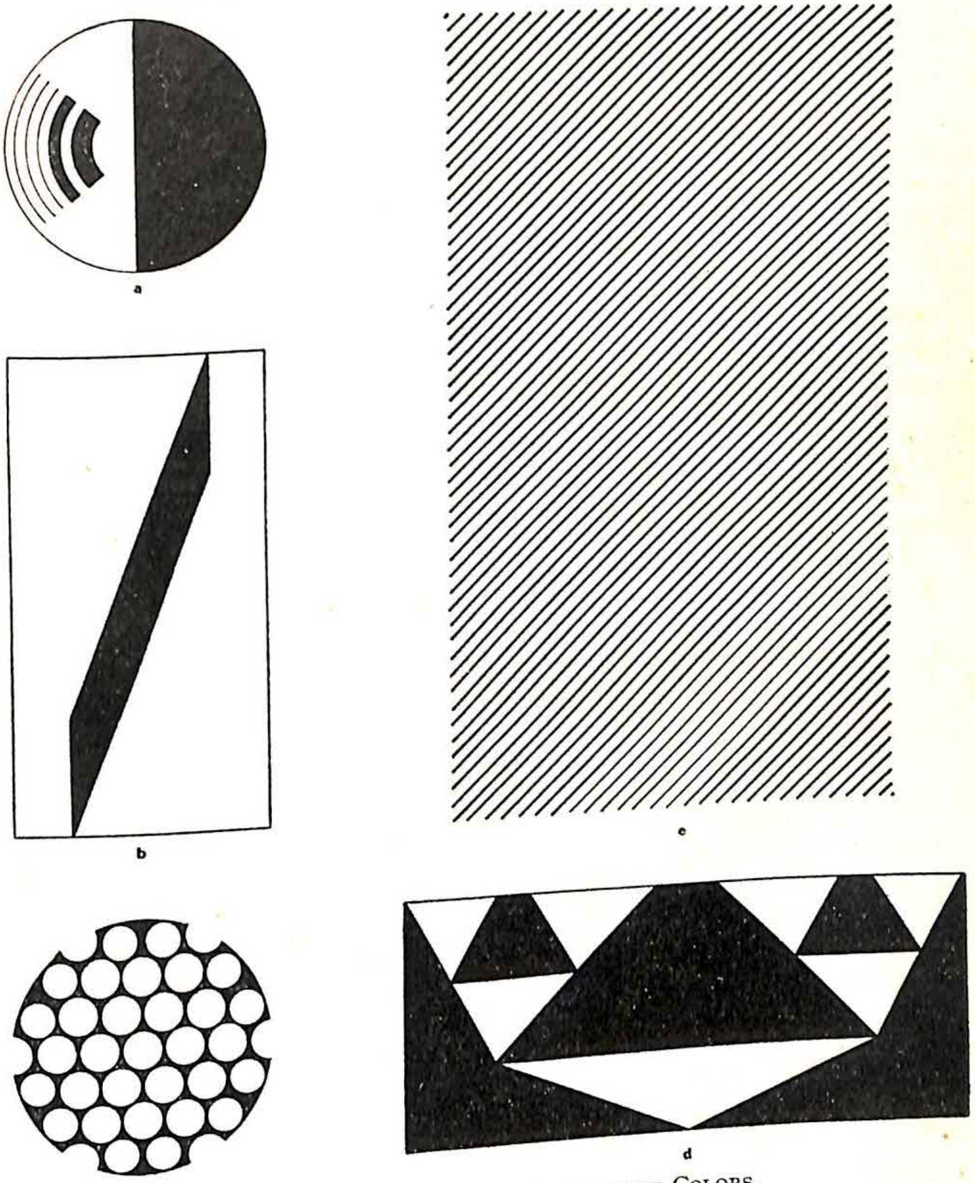
In the first paper (**46**), the assumption is made that the visual stimulus is transmitted through a layer of matter having a coefficient of diffusivity which is different for different wave-lengths, and varies too with the intensity of the stimulus. These theoretical assumptions are discussed in some detail.

> In order to investigate further this visual diffusivity of the eye, Ives constructed an apparatus consisting of two concentric circular disks of sheet metal, one outside the other, with a narrow radial slot on each. The relative position of these radial slots could be varied. Colored glasses could be placed in the slots. The speed of rotation of the disks could be varied. A source of illumination was placed behind the disk.
>
> When one slot was of a dark neutral tint while the other was of a lighter neutral tint, and the disk rotated, then the dark slot appeared several degrees behind the other. At very high or very low intensities, this phenomenon was less marked.
>
> When one slot was covered by a blue filter, and the other by a red filter, the blue image lags behind the red image by several degrees. If red, green, and blue filters are used, the red is first, then green, and then blue. Ives writes, "This lagging of one colour behind the other is a very striking phenomenon, well suited to demonstration to a small audience. The effect is most extraordinary if the disk, instead of being rotated, is rapidly oscillated through five or ten degrees, then the blue image may be made to vibrate entirely out of phase with the red, appearing as though attached to it by a cord."
>
> If a purple (mixed) color is revolved, the two components do not resolve (give one slot of red and one slot of blue). Ives explains this by saying, "the difference in diffusivities peculiar to two colors is much reduced when they are simultaneously transmitted to the same retinal area, although it is still present in some degree."

Nevertheless, the following year Ives (**47**) did succeed in resolving a color. He built a new apparatus, which utilized a projection lens system and a rotating mirror. He proposed to determine experimentally whether a mixed yellow resolved into its constituent red and green, and, whether a pure yellow will be resolved into anything at all except pure yellow. Ives found that the compound yellow was *clearly resolved* into red and green by a lateral motion across the field of view. Pure yellow did not separate.

Actually, Ives never did integrate his results with purple and yellow. We may point out that they were obtained with different apparatuses and that may have something to do with the experimental results. But what is important, and what Ives failed to realize, is that he did get a stimulus to resolve. The resolution is probably not caused by the differential diffusivity as Ives supposes, but by the differential rise and fall of the receptors (see **14**). Further work using his technique should yield most valuable results.

In 1922, DuBois (**28**), experimenting with mollusks, noticed that visual receptors respond at different rates to different wave-lengths.



a

b

e

d

c

FIG. 5. PATTERNS FOR THE PRODUCTION OF SUBJECTIVE COLORS.

Fig. 5a is Piéron's disk, similar to that of Doniselli. Fig. 5b is Kormann's design for use on a kymograph drum. Fig. 5c is Skinner's disk. The pattern is stationary and is viewed through a pinhole. Fig. 5d is Colenbrander's design for use on a kymograph drum. Fig. 5e is Luckiesh and Moss' design. The pattern is stationary. The colors are seen as pastel shades perpendicular to the parallel lines.

He proposed a theory closely analogous to the modulation theory of Troland (**78**). He observed, in passing, that the color experiences of a human observing the Benham disk may have a similar basis.

The first three of four important papers by Piéron (**58, 59, 60**) appeared in 1922 and 1923. Piéron states the problem of the subjective colors as being psychophysiological in nature. Physical theories such as those of Benham are impossible since subjective colors are seen in monochromatic light. The colors cannot be the same as after images because of the difference in time of the two phenomena. They are, he believes, related to such occurrences as the "fluttering heart," as Helmholtz had suggested. Again, he proposes the Fechnerian theory as a basis for the colors, although he does not give credit to Fechner. The curves, he states, of the rise and fall of the receptors ought to be deducible from the colors observed during the course of time. Further characteristics of the curves should be deducible from the fact that the curves rise faster when the intensity of the stimulating source is higher.

Piéron devised a Benham-type disk, after those of Maxwell, in which he could vary the speed of rotation, the black sector, the white sector, and the length of the lines. It is really an elaborate version of the apparatus of Doniselli (see Fig. 4a), and one form is illustrated in Fig. 5a.

The first series of experiments indicated that the diverse colors correspond to the defined delays (angles between black and beginnings of arcs) of the generating arc in relation to the beginning of the white sector, when the angle of the arc is kept constant.

The second series of experiments showed that the colors change under changes in illumination. But these changes can be compensated for so that the color will appear constant (but with perhaps a different saturation), if either or both the speed of rotation and/or the angular white section are changed. He shows that the characteristic delay of each color varies inversely as the fourth root of the intensity of the stimulus light. Experimentally, if the intensity is too high or too low, no color is seen.

The third series of experiments on the Benham disk indicated that thin lines show optimum subjective colors. As Finnegan and Moore previously noted, as the lines become thicker, color is seen only on the edge. If the lines become very thick, no color is seen at all.

Under normal illumination, the optimum length of the arc is about 45° (the lengths of the arcs of the Benham top). If the arc is 180°, no color is seen at all. Less color is seen at angular lengths between 180° and 45°, and less than 45°.

The fourth series of experiments showed that the amount of white sector following the lines has little influence on the colors seen. When the white precedes the lines, the amount of white determines the colors, with the exception of blue, which is always found in the terminal position. Several disks are produced in Piéron's paper which, although widely different, illustrate such equivalence of color. The shortening of the black sector when the speed of rotation is sufficiently rapid, weakens the blueing effect of the "privileged" terminal position of the lines. Under the same conditions the leading arc goes from red to green because the delay is longer. The elongation of the black sector has the opposite effect on both the leading and terminal lines.

The fifth series of experiments involved interpolating a segment of the black sector in the white sector, and on a different disk, the reverse procedure. Insertion of a small black sector in the middle of the larger white sector tends to blue the lines which precede, and redden those which come after. The interposition of a white phase on the black sector has an effect analogous to the simple shortening of the black sector, and the result is as described above.

In the sixth series of experiments Piéron compared the color appearances of the Benham disk under radiations through Wratten filters (running the gamut of the spectrum), with color appearances under white light, in much the same manner as Stewart and Abney. He finds that colors are seen which are not included in the "monochromatic" radiations, as did the previous investigators, but he is unable to derive any law of formation.

Finally, if a negative of the Benham top (similar to that of Finnegan and Moore, see Fig. 10) is rotated, colors are never seen on the white arcs, although colors are seen on other parts of the disk.

From his experimental results, Piéron attempted to deduce the curves of the rise and fall of the receptors. The curves, which change in form depending upon the illumination, are essentially like Stewart's third case (see Fig. 3c), with the difference that Piéron proposes a so-called "hypermaximal stage," *i.e.*, where each receptor responds to a high point before reaching an equilibrium. The receptors invariably rise in the order red, green, blue, and fall in the reverse order.

The black rings of the Benham disk appear to be colored by "diffusion" from the neighboring white surfaces which are being pulsed. The color seen on the lines is determined by the amount of activity in the lines and the resulting amount of irradiation. Thus, if there is no activity in the lines, *i.e.*, the lines follow the black directly, red is seen because of the strength of the diffusion. If there is some activity in the lines, *i.e.*, the lines follow white which follows the black, then some other color is seen because there is less diffusion.

In monochromatic light the color pulsed tends to stimulate and fatigue certain receptors, which allow the antagonistic receptors to discharge, and the complementary color is seen. Thus it follows that stimulation at first tends to influence the receptors most highly tuned the stimulus, and thereafter the complementary receptors. Piéron experimental results are quoted to prove this hypothesis.

Piéron concludes by formulating two laws which will predict colors of the Benham disk:

Law I. Under the influence of a luminous excitation of the retina by a complex radiation which integrates to make white there is first produced a ch matic disequilibrium with successive predominant shades running from red to blue in order of the spectral colors, because of an unequal speed of establishment of the fundamental chromatic processes released by this excitation until the hypermaximal transitory state is attained.

Law II. When a small retinal surface is not, or is only feebly excited, in the neighborhood of a region which is the site of a luminous and chromatic excitation, the excitation of this surface by diffusion shows a predominance of chromatic processes which are propagated with an intensity much greater than the luminous processes.

Forester (35) pretended to use the Benham disk in proof of his highly improbable theory of color vision. He proposes two levels of excitation in the nervous system relating to color perception. The first is in the retina, called the R level, and the second is in the optic tract, and is called the T level. White light has the interesting property of being able to by-pass the R level and influence only the T level. It nullifies the centers at the T level, and prohibits them from being influenced by the receptors at the R level. White is the physical stimulus on the rotating Benham disk (with the exception of the black sector) except in one place—where the lines are on the white surface. The red receptor can and does stimulate this spot, since the receptors are not nullified at the T level but why it should be red and not some other color seen is not made clear. The theory makes extensive use of the concept of Lapicque's chronaxie in its exposition.

Pauli and Wenzl (56) in 1924 also attempted an experimental analysis of the colors produced on the Benham disk. The experimental variables were the composition of the disk (distribution of black, white, and colored sectors), number of arcs, size of the segments and openings, diameter of the disk, color of the arcs, color of the dark sector, and speed and direction of rotation. These factors were varied by means of a Maxwell-type disk.

Pauli and Wenzl report that the colored appearances of the Benham disk can be divided into three groups based on the speed of rotation: the first at 3–5 rotations per second, the second at 5–37 rotations per second, and the third at 33–40 rotations per second. The colored appearances are given in great detail and will not be discussed here—except to mention that in the second phase a mosaic pattern is seen, which may be the hexagonal pattern seen by others.

If the white field is replaced by a colored field, after the manner of Bagley, the black sector shows complementary colors. If the black field is replaced by a colored sector, the colors are still seen, but they may be altered somewhat, depending on the particular color field substituted.

Subjective colors may be seen during tachistoscopic exposure.

Under these conditions the colors show a decrease in intensity and distinctness, and can be seen only at a fast speed of rotation. In direct sunlight many of the colors disappear completely. When the disk is viewed peripherally, no colors may be seen. In general, Pauli and Wenzl's other results agree with former investigators.

Pauli and Wenzl, in their theoretical explanation, state that after images are not responsible for the colors. They theorize that each of the receptors has a typical inertia, and consequently a typical natural period. When the frequency of the pulse is that of the natural period of the receptors, it will respond. Pauli and Wenzel, after the manner of Piéron, also postulate an induction phenomenon in explaining the spread of the colors. "We have," they say, "an overlaying of three factors; primary excitation, resonance and number of excitations, and the induction effect."

Von Kries (81), writing in the 1924 American translation of Helmholtz, pointed out that Helmholtz' earlier description of the subjective colors could be applied to the Benham top.

Stewart (77) in 1924 wrote his second paper on the subjective colors. Observations were made by looking at a rotating mirror which reflected the light source. The colors appeared as luminous bands, and, in the main, the color appeared homogeneous over the entire surface except at the edges. However, the color phenomena on the edges of the band were essentially the same as the rest of the band, but different in phase, because the edges were illuminated by slightly less intensity.

In general, a white light of given intensity may be pulsed at progressively increased speeds so that the subjective color sensation passes progressively from violet to red approximately in the ordering of the spectrum. The colors after the green on the red end of the spectrum are less pure than the others because the pulse rate is rapidly approaching the fusion point. However, under suitable conditions, when the illumination is carefully controlled, a very good red can be obtained. It can further be shown that when the pulse rate is held constant, the same succession of colors can be obtained by adjusting the intensity of the light; when the physical intensity is increased the colors shift toward the violet and when the physical intensity is decreased, the colors shift toward the red.

Color effects can be observed by looking at a black and white picture through a black, white, and sectored disk (similar to Bidwell, Fig. 1u). The color effects are most determined by the physical intensity of the illumination of any given portion of the picture, in other words, the shading. Stewart writes:

> It is easy to cause the darker foliage in an engraving of a landscape to assume a green color while the less shaded water or sky is violet or bluish, with the details of the shading represented by corresponding nuances of the subjective color. Then by a change of speed of the disk a complete change can be produced in the colors, which again, however, are appropriately placed on the different parts of the picture according to the depth of the shading. The colors are so easy to obtain that I have seen children amuse themselves by the hour "coloring" pictures in books or newspapers subjectively by observing them through a rotating disk, and describing the colors correctly. While it is generally not difficult to clothe the landscapes in colors which are appropriate, or at least, not glaringly incongruous, it is of course different with pictures in which human faces or figures are the important element. With the shading below the chin perhaps a light green (at a particular speed of rotation), the cheeks violet or bluish, the hollows of the eyes perhaps green, the hair a dark green, and various garments colored according to their shading (creases, for instance, at a given phase green), and then everything changing phase with an increase or diminution in the speed, the personal appearance of a state beauty, the statesman or criminal of the hour, the horse which has just won the Derby, or the reigning pugilist, is apt to be somewhat bizarre, and therefore rather amusing for children. On the other hand black and white pictures of such objects as embroidery, tapestry, Scotch tartans, carved work, stained and cut glass, etc., may show curious and even beautiful appearances, the subjective colors imparting a certain richness to the designs. Of course, unless the black and white pictures were expressly produced in order that the subjective tints might approximate the actual colors at some rate of intermission (a feat perhaps in simple cases theoretically possible), the subjective colors could only by accident be correct.

Stewart includes a rather elaborate discussion of his observations of a rotating Benham top. In general his results agree with those of previous investigators. When the lines are observed through a magnifying glass, the color is the same as when observed with the naked eye. When the Benham top is held stationary and observed through a rotating sectored disk, the color appearances on the Benham top are about the same as those seen when the top itself is rotating. Of course, all the arcs appear to be the same color, but the color can be made to change depending on the speed of rotation of the sectored disk. In experiments utilizing Benham disks with arcs of varying thickness, colors can be seen throughout the arc no matter what the thickness, although different colors are seen in different portions—a function of the speed of rotation. Color blind individuals do not see the subjective colors with the same clarity as normal individuals.

Stewart conducted experiments wherein colored lights were observed in the rotating mirror and through the rotating disk. He also viewed colored pictures under the same conditions. He reports that with a slow pulse rate, the blue tints are strengthened, white becomes violet and the greens and reds are weakened. As the pulse rate is increased, the blue remains predominant and the greens are strengthened. At a high rotational speed, the reds prevail. The color changes in the picture are primarily dependent on the physical saturation, hue, and brightness of the picture; the Purkinje phenomenon is not a factor.

Concerning the effect of the subjective color sensations on evolution, he writes:

> The question suggests itself whether the color sensations of animals in an environment habitually or frequently illuminated by flickering light, as in forests or not far below water surfaces, could be modified at all by such phenomena as those under discussion. If so, protective color adaptation might in some degree be expected under these circumstances to take account of the factor of subjective coloration. Protective coloration has considerable interest in connection with the question of color vision in animals. For it seems in some cases to afford a means of testing whether the response in sensation to given wave-lengths is approximately the same as in man. Whether the subjective modifications of color sensation caused by intermission at appropriate rates enter at all as a factor into such phenomena as the aurora borealis, or some other of the solar color phenomena, depends upon the magnitude of any intermittent changes of luminosity concerned in them and their period. Where the spectral characters of the light sufficiently explain the perceived colors there is of course no room left for the subjective factors.

Piéron (61) published his fourth paper on the subjective colors in 1928. Essentially, it was a modification of the work of Abney. He observed in a most systematic fashion a rotating Benham top under various white lights. These white lights were always composed of two near monochromatic radiations. The whites looked alike; their spectral compositions were vastly different. Finally, the top was observed under a yellow source, composed of red and green radiations.

Piéron's results conclusively show that the color seen is a function of the spectral composition of the illuminating source and not of its tristimulus values. The results give rise to two principles. (1) The dominant radiations of the spectral composition of the illuminating source tend to become accentuated. (2) The complement of the dominant radiations tends to be evoked. The second principle may be in opposition to the first in that it may create a color which corresponds to absent radiations.

Kormann (50) in 1931 described a "new" type disk which revolved on a kymograph drum. He was apparently unaware of the similar work performed by Doniselli and Piéron. He presents three new designs, one of which is illustrated in Fig. 5b. He points out that if the design were transferred to a disk, the result would be similar to a half moon—like Wolf's disk (see Fig. 1t) which he does not mention. Kormann compared his artificial spectrum with a prismatic spectrum and reported

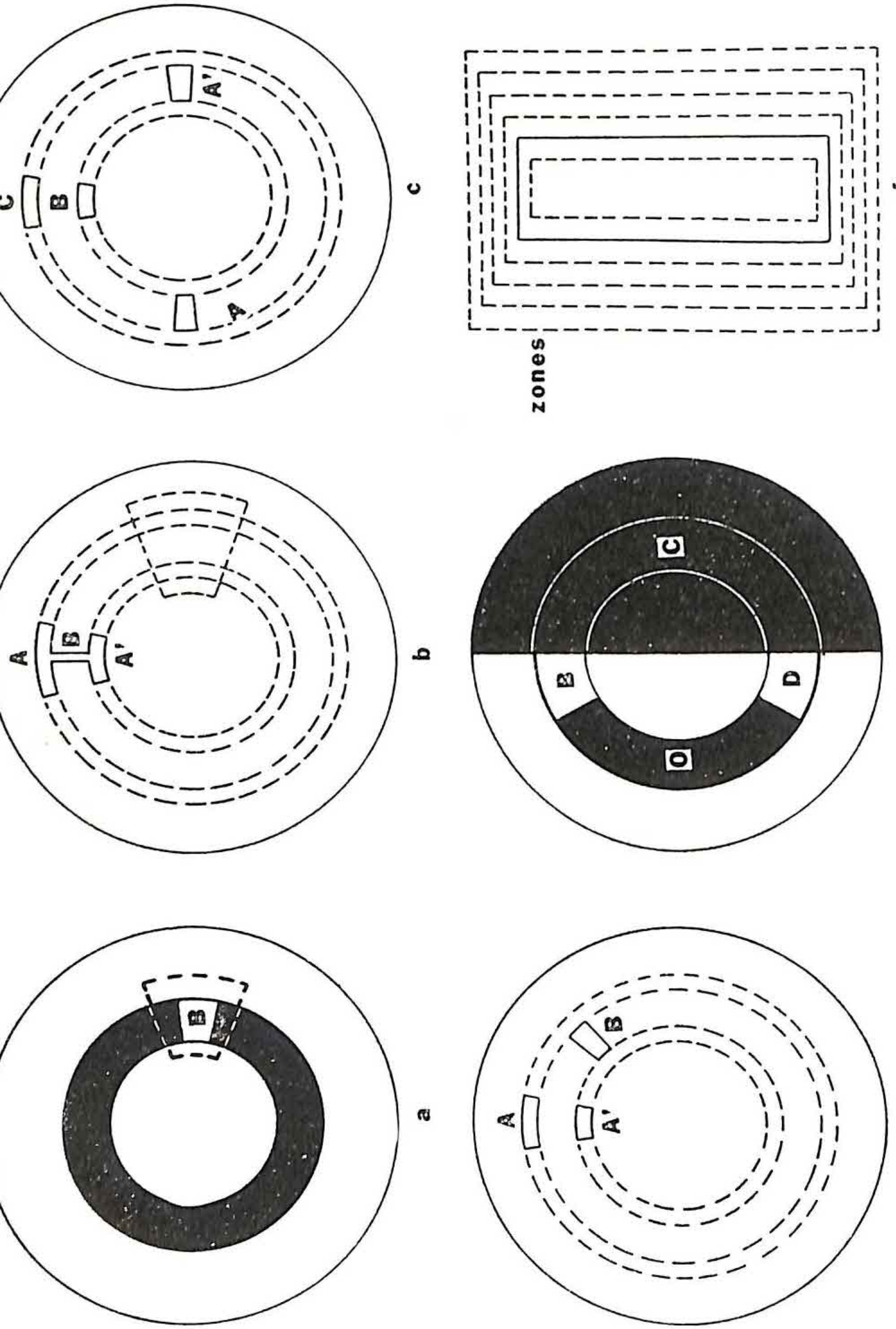


FIG. 6. THE SPECIALIZED CHARACTERISTICS OF THE SUBJECTIVE COLORS.
Figs. 6a–6e are Fry's disks illustrating the effects of interaction. Figs. 6a–6d are sectored. Fig. 6f depicts Gebhard's zoning effects.

given intensity of stimulation increasing the duration causes the color to change progressively in the direction from green toward pink. . . Beyond a certain limit, however, increasing the duration causes the color to change to white.

For the final experiment in this paper, the surface illumination of the disk was made both brighter and equal to *B*. Pink can be produced without the aid of bright surroundings, while purple, blue, and green cannot.

All of these experimental results, some of which were known before, are exceptionally interesting and provocative. The fact that they do not disprove the modulatory theory is only of minor importance since they do not disprove a number of other theories, for example, that of Forster (35). Actually, in science, we *never* prove that a theory is right, but only that it is not wrong. Fry's analysis in terms of the modulation theory would have been a more powerful document had it shown the Fechnerian theory to be wrong.

Fry's (37) second paper in 1933 makes inquiry concerning the stimulated retinal area, beta, and its surround. If beta and the surround can be pulsed or not pulsed, there are four possible combinations: beta is steady, surround steady; beta steady, surround intermittent; beta intermittent, surround steady; beta intermittent, surround intermittent. If it can be shown that the color seen is dependent solely on the intermittence of beta, then no other facts need be invoked for an explanation. The color seen in beta can be considered to be a color contrast phenomenon if the color is independent of the pulsing of beta and dependent on the pulsing of the surround. The level of the intensity of the surround, whether it pulses or not, can be shown to affect the colors seen in beta.

Fig. 6b illustrates Fry's disk for this series of experiments. *A*, *B*, and *A'* indicate slits. The dotted lines indicate an illuminated background. When *A*, *B*, and *A'* are all intermittent, *A* and *A'* appear pink and *B* appears green. When *A* and *A'* are steady and *B* is intermittent, then *A* and *A'* appear pink and *B* appears green. When *A* and *A'* are pulsed, and *B* is intermittent, then *A* and *A'* appear pink, and *B* still appears green. The contrast effect is non-operative. Fig. 6c indicates an additional stimulus pattern. *A*, *A'*, *B* and *C* are slits. When the disk is rotated, *b* is seen as yellow. If *b* remains steady, and *A* and *A'* are intermittent, *b* still appears yellow. If *A* and *A'* are steady, then *b* appears blue. The yellow of *b* is a contrast phenomenon. Similarly, the contrast effect by *b* is responsible for the colors seen in *A* and *A'*. Fig. 6d indicates another stimulus pattern illustrating interaction. When rotated in a counterclockwise direction, *A* and *A'* appear purplish blue, and *B* appears yellowish white. In a reverse direction, the colors are reversed.

Fry concludes this paper with a discussion of the Benham top, adding that the red seen is due, at least in part, to chromatic aberration.

On a different disk, subjective red borders, when viewed through a convex lens, appear to be blue.

Fry takes issue with Piéron's concept that the leading white sector is the determiner of the color. He observes that a disk shown in Fig. 6e, when rotated, will shown green. It will still show green no matter how much the inner circle *CBDO* is rotated with respect to the rest of the disk. It is only when the absolute amounts of white at *B* and *D* are changed that the color is changed.

Fry's (**38**) third paper, published in 1935, attempted to show that the trichromatic theory (our Fechnerian theory) was inadequate in its explanation of the subjective colors. He constructed a rather elaborate apparatus which presented the same spectral source to two fields, which could be observed by the subject. The intensity of one field could be varied by adjustment of the sector slit of an episcotister, while the intensity of the other could be varied by adjustment of the speed of rotation of another episcotister. The conditions of equivalence between the two fields were studied for different brightnesses of ground and for different wave-lengths. That the ground is important is illustrated by a series of curves, but their theoretical implications need not concern us here.

A more important series of curves is that which designates equivalent brightnesses between the two fields when wavelength is varied. According to Fry, the fact that the red response is relatively weak finds basis in how one interprets the response of the receptors. If the red response is due primarily to the response of the red receptor, then we can say that the red receptor is weak. Experimentally, he demonstrates that a bluish green sensation can be obtained from white light which contains all the colors of the spectrum, and also from white which is composed only of blue and green. This may give support to the concept of the feeble response of the red. Three facts remain unexplained. Colors other than blue green can be obtained with an intermittent white light. If the stimulating light is of sufficient intensity, or the exposures are long enough, or both, then a pink sensation results. The blue green sensation cannot be produced unless the surround is of high intensity. Fry concludes, "As in the case of the eye, so also in the case of the photographic plate (**15**), intermittent exposure is less effective than continuous exposure for all wavelengths, and the longer waves suffer more than the rest in this respect." The three component theory, says Fry, offers no ready explanation for these facts.

Fry's (**39**) fourth paper, published in 1936, makes use of an apparatus similar to that described in his third paper. Fry investigated the subjective colors seen under pulsing monochromatic lights, and under pulsing monochromatic lights whose saturations were reduced by a non-pulsing white light. The effects of varying the wavelength and frequency of the pulse were investigated. The experiments are similar to those of Abney and Piéron, to which Fry makes no mention. In general, the color seen is that of the hue of the wave-length of the stimulating light. There is, however, a tinge of blue and violet, in the secondary images, thought to be similar in nature to Bidwell's ghost. When stimuli of lower saturation (as described above) are pulsed, the secondary images tend to be complementary to the primary stimulus.

Erb and Dallenbach (31) presented a most important contribution in 1939. They proposed to study the Luckiesh-Moss stimulus pattern (see Fig. 5e) to determine the effect of knowledge and instruction, color of the pattern and color of the background, the width of lines and the spacings, and eye movements on the subjective colors.

> Observers with and without knowledge of the subjective colors were shown the stimulus pattern with three different types of instructions. One group was told to simply observe the pattern, the second group was told to use far fixation, and the third group was told to look for colors. Knowledge of the phenomenon favors the subjective colors, but it by no means guarantees that they will be seen. All observers told to look for the colors saw them. Phenomenal movement was reported by very many subjects: streaming, random movement, moving band and zig-zag lines, movement of stimulus lines, circular movement, and shimmer. There is no relation between the subjective colors and the subjective movement.
>
> The pattern was printed in red, yellow, green, blue, black, and white upon red, yellow, green, blue, black, neutral gray, and white backgrounds. It was also printed upon various stocks of paper. The following combinations were found to give the optimum colors: black lines on white background, blue on white, green on white, black on green, black on yellow, and green on yellow. The best of these is black on white.
>
> Narrow lines and narrow spaces give saturated hues; narrow lines and wide spaces, or the reverse, give less saturated colors. Broad lines and wide spaces give no color at all.
>
> When the stimulus pattern is placed in the Dodge tachistoscope, and the exposure is very short, no colors are seen. An exposure of very much greater length than that required for eye movements is necessary before the subjective colors are observed. It follows, that although eye movements are a necessary factor, they are not sufficient or compulsory.

Abel (1), a student of Dallenbach, continued these researches in 1939. Her problem was to observe the subjective colors of the same stimulus pattern under monochromatic and polychromatic illumination. The following light sources were used: a tungsten lamp radiating a continuous spectrum; a mercury arc lamp with its lines at 577, 579, 546, 492, 436, and 405 millimicrons; an amber filter which transmitted almost exclusively the 577 and 579 millimicron lines from the mercury arc; a green filter which transmitted the 546 line of the mercury arc; and a red filter, used with the continuous source, with a sharp cut off at 623 millimicrons. Abel writes:

The main results may be classified in the following manner: (1) Differentiation among the black diagonal lines both of hue and of brightness. (2) Colors in the interspaces between the black lines either filling in all the space between the lines or merely along the edges of the lines. (3) Pattern lines diagonally across the parallel black lines. Sometimes these patterns were black or gray but more often colored pastel shades. (4) When colored phenomena were reported, continual change of the color was observed generally in the direction of an increase in saturation.

Under all conditions, both with polychromatic and monochromatic lights, with the exception of monochromatic red, many subjective colors were reported. In polychromatic light, the colors were usually more saturated than they appeared in monochromatic light. The colors seen were not necessarily restricted to the after image of the physical color. With red light, only various shades of red are seen. These results are in direct opposition to those reported by Fry.

Segal (67), a student of Piéron, in 1940, independently hypothesized the modulation theory of Troland, but did not mention either Troland or Fry. He too points out that the subjective color appearances are not incongruous with the theory.

Gebhard (40) contributed the most recent, thorough communication in 1943. He utilized the left side of an apparatus described by Brown (17). The subject observes a figure and ground. For both this figure and ground, it was possible to vary the luminosities independently, the frequencies of intermittence, the on/off ratios, and the phase relationships. Thus was allowed a great variety of experimental conditions. Three kinds of illumination were used: a white continuous source, a Wratten red filter which transmitted nothing below 650 millimicrons, and a filter combination for wave-lengths at 560 millimicrons.

Observations were made with white, red, and green lights, employing various frequencies and conditions whereby the ground had a phase lead over the field, and vice versa. For two of the three observers, the red and green stimuli produced no subjective sensations other than those attributable to the physical characteristics of the stimulus. The third observer saw a multitude of colors. It is possible, Gebhard suggests, that the difference was due to the observer's greater astigmatism. This led to the second problem.

The boundary between the field and ground was made indistinct. Under these conditions all observers saw colors—some where they had not seen them before. The field was now made into a rectangular bar whose width could be varied. Now, "the area of the figure and the part of the field immediately adjacent to it, was found to divide itself symmetrically into a series of zones or bands parallel to the two lateral boundaries of the bars." This effect is illustrated in Fig. 6f. The wider the bars, the greater the number of zones which are produced. When the illumination was reduced, two of the three subjects could not see zones. Practice and fatigue are necessary for the optimum perception of the zones. Simultaneous with the zoning effect, there were many

field color effects. Geometrical patterns are mentioned again, and these may be the same as Fechner's hexagonal cells.

Gebhard attempts to show the inadequacy of the Fechnerian theory by saying, "Colors are producible under various conditions of stimulation with monochromatic red illumination so chosen as to act upon the R retinal process only." One wonders why Gebhard assumes that he is stimulating one retinal process with a filter transmitting 650 millimicrons and above. The further inadequacy is shown by, "The slow arousal of the zoning effect with blurring and the general instability reflected in long term fluctuation of the zones." Gebhard continues, ". . . it is doubtful if the modulation hypothesis is sufficiently comprehensive to deal with the formation of multiple boundaries, the intermittent and steady field effects, the instability of the colors and zones and the long term developmental effects."

Finally, Brown and Gebhard (**18**) in 1948 reported several new color phenomena, some which were of a binocular nature, which may be related to the subjective colors.

The present paper has made no attempt to cover the literature on the direct measurement of the rise and fall of the color receptors. An excellent survey of this material may be found in Bills (**14**).

VIII. SUMMARY

1. It is suggested that the Prevost-Fechner-Benham subjective colors may be used as an outside criterion for the determination of the axes in the color space.

2. The subjective colors were discovered by Prevost (**62**) in 1826. They were rediscovered by Fechner (**33**), John Smith (**73, 74**), Brewster (**16**), F. J. Smith (**72**), Hannay (**42**), Stewart (**76**), Charpentier (**21**), Benham (**84**), Ives (**46, 47**), Klopsteg (**49**), Way (**82**), and Skinner (**71**)

3. Subjective colors may be seen in any white light which can be made to pulse below the fusion frequency. Such colors have been reported by Prevost (**62**), J. Smith (**73, 74**), F. J. Smith (**72**), Hannay (**42**), Stewart (**76, 77**), Charpentier (**21, 22, 23**), Bidwell (**11**), Percival (**57**), and Fry (**39, 40**).

The colors are most easily seen on rotating black and white disks of various designs. Such disks (or designs on kymograph drums) have been presented by Fechner (**33**), Dove (**26, 27**), John Smith (**73, 74**), Helmholtz (**43**), Brücke (**19**), Benham (**84**), Finnegan and Moore (**34**), Hurst (**45**), Wolf (**85**), Baumann (**6**), Doniselli (**25**), Edridge-Green (**30**), Kormann (**50**), and Colenbrander (**24**).

Colors have been obtained from sectored disks by Rood (**63**), Brewster (**16**), *La Nature* (**41**), Bidwell (**11**), and Fry (**36, 37**).

The main results may be classified in the following manner: (1) Differentiation among the black diagonal lines both of hue and of brightness. (2) Colors in the interspaces between the black lines either filling in all the space between the lines or merely along the edges of the lines. (3) Pattern lines diagonally across the parallel black lines. Sometimes these patterns were black or gray but more often colored pastel shades. (4) When colored phenomena were reported, continual change of the color was observed generally in the direction of an increase in saturation.

Under all conditions, both with polychromatic and monochromatic lights, with the exception of monochromatic red, many subjective colors were reported. In polychromatic light, the colors were usually more saturated than they appeared in monochromatic light. The colors seen were not necessarily restricted to the after image of the physical color. With red light, only various shades of red are seen. These results are in direct opposition to those reported by Fry.

Segal (**67**), a student of Piéron, in 1940, independently hypothesized the modulation theory of Troland, but did not mention either Troland or Fry. He too points out that the subjective color appearances are not incongruous with the theory.

Gebhard (**40**) contributed the most recent, thorough communication in 1943. He utilized the left side of an apparatus described by Brown (17). The subject observes a figure and ground. For both this figure and ground, it was possible to vary the luminosities independently, the frequencies of intermittence, the on/off ratios, and the phase relationships. Thus was allowed a great variety of experimental conditions. Three kinds of illumination were used: a white continuous source, a Wratten red filter which transmitted nothing below 650 millimicrons, and a filter combination for wave-lengths at 560 millimicrons.

Observations were made with white, red, and green lights, employing various frequencies and conditions whereby the ground had a phase lead over the field, and vice versa. For two of the three observers, the red and green stimuli produced no subjective sensations other than those attributable to the physical characteristics of the stimulus. The third observer saw a multitude of colors. It is possible, Gebhard suggests, that the difference was due to the observer's greater astigmatism. This led to the second problem.

The boundary between the field and ground was made indistinct. Under these conditions all observers saw colors—some where they had not seen them before. The field was now made into a rectangular bar whose width could be varied. Now, "the area of the figure and the part of the field immediately adjacent to it, was found to divide itself symmetrically into a series of zones or bands parallel to the two lateral boundaries of the bars." This effect is illustrated in Fig. 6f. The wider the bars, the greater the number of zones which are produced. When the illumination was reduced, two of the three subjects could not see zones. Practice and fatigue are necessary for the optimum perception of the zones. Simultaneous with the zoning effect, there were many

field color effects. Geometrical patterns are mentioned again, and these may be the same as Fechner's hexagonal cells.

Gebhard attempts to show the inadequacy of the Fechnerian theory by saying, "Colors are producible under various conditions of stimulation with monochromatic red illumination so chosen as to act upon the R retinal process only." One wonders why Gebhard assumes that he is stimulating one retinal process with a filter transmitting 650 millimicrons and above. The further inadequacy is shown by, "The slow arousal of the zoning effect with blurring and the general instability reflected in long term fluctuation of the zones." Gebhard continues, ". . . it is doubtful if the modulation hypothesis is sufficiently comprehensive to deal with the formation of multiple boundaries, the intermittent and steady field effects, the instability of the colors and zones and the long term developmental effects."

Finally, Brown and Gebhard (**18**) in 1948 reported several new color phenomena, some which were of a binocular nature, which may be related to the subjective colors.

The present paper has made no attempt to cover the literature on the direct measurement of the rise and fall of the color receptors. An excellent survey of this material may be found in Bills (**14**).

VIII. SUMMARY

1. It is suggested that the Prevost-Fechner-Benham subjective colors may be used as an outside criterion for the determination of the axes in the color space.

2. The subjective colors were discovered by Prevost (**62**) in 1826. They were rediscovered by Fechner (**33**), John Smith (**73, 74**), Brewster (**16**), F. J. Smith (**72**), Hannay (**42**), Stewart (**76**), Charpentier (**21**), Benham (**84**), Ives (**46, 47**), Klopsteg (**49**), Way (**82**), and Skinner (**71**)

3. Subjective colors may be seen in any white light which can be made to pulse below the fusion frequency. Such colors have been reported by Prevost (**62**), J. Smith (**73, 74**), F. J. Smith (**72**), Hannay (**42**), Stewart (**76, 77**), Charpentier (**21, 22, 23**), Bidwell (**11**), Percival (**57**), and Fry (**39, 40**).

The colors are most easily seen on rotating black and white disks of various designs. Such disks (or designs on kymograph drums) have been presented by Fechner (**33**), Dove (**26, 27**), John Smith (**73, 74**), Helmholtz (**43**), Brücke (**19**), Benham (**84**), Finnegan and Moore (**34**), Hurst (**45**), Wolf (**85**), Baumann (**6**), Doniselli (**25**), Edridge-Green (**30**), Kormann (**50**), and Colenbrander (**24**).

Colors have been obtained from sectored disks by Rood (**63**), Brewster (**16**), *La Nature* (**41**), Bidwell (**11**), and Fry (**36, 37**).

The main results may be classified in the following manner: (1) Differentiation among the black diagonal lines both of hue and of brightness. (2) Colors in the interspaces between the black lines either filling in all the space between the lines or merely along the edges of the lines. (3) Pattern lines diagonally across the parallel black lines. Sometimes these patterns were black or gray but more often colored pastel shades. (4) When colored phenomena were reported, continual change of the color was observed generally in the direction of an increase in saturation.

Under all conditions, both with polychromatic and monochromatic lights, with the exception of monochromatic red, many subjective colors were reported. In polychromatic light, the colors were usually more saturated than they appeared in monochromatic light. The colors seen were not necessarily restricted to the after image of the physical color. With red light, only various shades of red are seen. These results are in direct opposition to those reported by Fry.

Segal (**67**), a student of Piéron, in 1940, independently hypothesized the modulation theory of Troland, but did not mention either Troland or Fry. He too points out that the subjective color appearances are not incongruous with the theory.

Gebhard (**40**) contributed the most recent, thorough communication in 1943. He utilized the left side of an apparatus described by Brown (**17**). The subject observes a figure and ground. For both this figure and ground, it was possible to vary the luminosities independently, the frequencies of intermittence, the on/off ratios, and the phase relationships. Thus was allowed a great variety of experimental conditions. Three kinds of illumination were used: a white continuous source, a Wratten red filter which transmitted nothing below 650 millimicrons, and a filter combination for wave-lengths at 560 millimicrons.

Observations were made with white, red, and green lights, employing various frequencies and conditions whereby the ground had a phase lead over the field, and vice versa. For two of the three observers, the red and green stimuli produced no subjective sensations other than those attributable to the physical characteristics of the stimulus. The third observer saw a multitude of colors. It is possible, Gebhard suggests, that the difference was due to the observer's greater astigmatism. This led to the second problem.

The boundary between the field and ground was made indistinct. Under these conditions all observers saw colors—some where they had not seen them before. The field was now made into a rectangular bar whose width could be varied. Now, "the area of the figure and the part of the field immediately adjacent to it, was found to divide itself symmetrically into a series of zones or bands parallel to the two lateral boundaries of the bars." This effect is illustrated in Fig. 6f. The wider the bars, the greater the number of zones which are produced. When the illumination was reduced, two of the three subjects could not see zones. Practice and fatigue are necessary for the optimum perception of the zones. Simultaneous with the zoning effect, there were many

field color effects. Geometrical patterns are mentioned again, and these may be the same as Fechner's hexagonal cells.

Gebhard attempts to show the inadequacy of the Fechnerian theory by saying, "Colors are producible under various conditions of stimulation with monochromatic red illumination so chosen as to act upon the R retinal process only." One wonders why Gebhard assumes that he is stimulating one retinal process with a filter transmitting 650 millimicrons and above. The further inadequacy is shown by, "The slow arousal of the zoning effect with blurring and the general instability reflected in long term fluctuation of the zones." Gebhard continues, ". . . it is doubtful if the modulation hypothesis is sufficiently comprehensive to deal with the formation of multiple boundaries, the intermittent and steady field effects, the instability of the colors and zones and the long term developmental effects."

Finally, Brown and Gebhard (**18**) in 1948 reported several new color phenomena, some which were of a binocular nature, which may be related to the subjective colors.

The present paper has made no attempt to cover the literature on the direct measurement of the rise and fall of the color receptors. An excellent survey of this material may be found in Bills (**14**).

VIII. SUMMARY

1. It is suggested that the Prevost-Fechner-Benham subjective colors may be used as an outside criterion for the determination of the axes in the color space.

2. The subjective colors were discovered by Prevost (**62**) in 1826. They were rediscovered by Fechner (**33**), John Smith (**73, 74**), Brewster (**16**), F. J. Smith (**72**), Hannay (**42**), Stewart (**76**), Charpentier (**21**), Benham (**84**), Ives (**46, 47**), Klopsteg (**49**), Way (**82**), and Skinner (**71**)

3. Subjective colors may be seen in any white light which can be made to pulse below the fusion frequency. Such colors have been reported by Prevost (**62**), J. Smith (**73, 74**), F. J. Smith (**72**), Hannay (**42**), Stewart (**76, 77**), Charpentier (**21, 22, 23**), Bidwell (**11**), Percival (**57**), and Fry (**39, 40**).

The colors are most easily seen on rotating black and white disks of various designs. Such disks (or designs on kymograph drums) have been presented by Fechner (**33**), Dove (**26, 27**), John Smith (**73, 74**), Helmholtz (**43**), Brücke (**19**), Benham (**84**), Finnegan and Moore (**34**), Hurst (**45**), Wolf (**85**), Baumann (**6**), Doniselli (**25**), Edridge-Green (**30**), Kormann (**50**), and Colenbrander (**24**).

Colors have been obtained from sectored disks by Rood (**63**), Brewster (**16**), *La Nature* (**41**), Bidwell (**11**), and Fry (**36, 37**).

Colors from black and white pictures, when viewed under a pulsing light, have been reported by Stewart (77).

Colors from stationary patterns, due to eye movements, have been reported by Skinner (71) and Luckiesh and Moss (54).

4. Almost all investigators report that the type and level of illumination, the speed of rotation or frequency of pulses, the design or distribution of black, white, and lines, and the state of the eye affect the colors seen.

Movement of some kind appears to be a necessary factor. That brief tachistoscopic exposure shows no colors at all is reported by Pauli and Wenzl (56) and Erb and Dallenbach (31).

That practice aids in the perception of the colors is reported by Helmholtz (43), Aubert (4), Gebhard (40), but Bagley (5) insists that practice is not important. Erb and Dallenbach (31) show that knowledge of the colors facilitates their observation.

That there are differences in the colors seen in direct and indirect vision is reported by Aubert (4), Exner (32), and Pauli and Wenzl (56).

Stewart (77) reports that color blind individuals do not see the colors with the same clarity as normal individuals.

Induction phenomena (subjective colors induced by the stimulation of an adjacent retinal area) are reported by Bagley (5), Piéron (58, 59, 60, 61), Pauli and Wenzl (56), Fry (37), and Gebhard (40).

5. A hexagonal cell structure, seen with the subjective colors, has been reported by Fechner (33), Brewster (16), Helmholtz (43), Aubert (4), Exner (32), Edridge-Green (29), and Gebhard (40).

6. Subjective colors from non-white stimuli fall into three classes. Those who reported subjective colors from general chromatic stimuli are Dove (26, 27), Rood (63), Helmholtz (43), Hannay (42), Stewart (76, 77), Benham (9), Liveing (53), Ives (46, 47), Piéron (58, 59, 60, 61), Pauli and Wenzl (56), Klopsteg (49), Colenbrander (24), Erb and Dallenbach (31), and Gebhard (40). Those who reported subjective colors from general chromatic stimuli with the exception of red are Abney (2) and Abel (1). Fry (39) reported that he was able to get no subjective colors from chromatic stimuli, except those of after images.

7. The theory of the differential rise and fall of the receptors was proposed by Fechner (33), and supported by Helmholtz (43), Hannay (42), Stewart (76, 77), Liveing (52), Sexton (68), Percival (57), Piéron (58, 59, 60, 61), and Colenbrander (24).

The modulation theory of the optic nerve is supported by DuBois (28), Troland (78), Adrian and Matthews (3), Fry (36, 37, 38, 39), and Segal (67).

Other theories of a physiological nature have been suggested by Prevost (62), N. Smith (75), Charpentier (22), Gray (86), *La Nature* (41), Henry (44), Bidwell (11), Bagley (5), Doniselli (25), Baumann (6, 7, 8), Edridge-Green (30), Ives (46, 47), Piéron (58, 59, 60, 61), Forster (35), Pauli and Wenzl (56), and Colenbrander (24).

Other theories of a physical nature have been proposed by J. Smith (73, 74), F. J. Smith (72), Benham (9), and Schwartze (66).

Discussion

This paper has outlined the long history of subjective colors. It may be that one day these same colors may solve the great enigma of color vision—the specification of the receptors of the eye.

Throughout the history of physiological optics the eye has never resolved a stimulus into its components under any other experimental conditions. The receptors could never be determined directly because the eye consistently integrates the many wave-lengths of the stimulus to perceive it as an unanalyzed sensation. This integration must require the most elaborate mechanism—since it fulfills this ideal even under the most elaborate experimental conditions, with but a single exception.

But it is just possible that the eye can be made to yield its secret. If the subjective colors are due to the differential rise and fall of the receptors, then here at last the eye is not so perfect after all, for it is not completely integrating the stimulating wave-lengths. If the Fechnerian theory of the subjective colors is true, then we need but explore the color space until we find colors which when pulsed, show no subjective colors. When such colors are found, then we are *stimulating one receptor*. If it be that the receptors are imaginary, as is probably the case, then data may be taken on the real colors, and extrapolated into the color space which corresponds to no physical stimulus. Of course, the experimental variables affecting the subjective colors, which have been discussed at length in this paper, would have to be carefully controlled.

BIBLIOGRAPHY

1. Abel, T. M. 'Subjective' colors from line-patterns viewed in polychromatic and monochromatic light. *Amer. J. Psychol.*, 1939, **52**, 610–615.
2. Abney, W. de W. The artificial spectrum top. *Nature, Lond.*, 1894, **51**, 292.
3. Adrian, E. D., & Matthews, R. The action of light on the eye. *J. Physiol.*, 1927, **63**, 378–414; 1927, **64**, 279–301; 1928, **65**, 273–298.
4. Aubert, H. *Physiologie der Netzhaut.* Breslau: E. Morgenstern für Aug. Schultz und Co., 1865. P. 355 and pp. 377–380.
5. Bagley, F. W. An investigation of Fechner's colors. *Amer. J. Psychol.*, 1902, **13**, 488–525.
6. Baumann, C. Beiträge zur Physiologie des Sehens: IV. Subjektive Farbenerscheinungen. *Pflüg. Arch. ges. Physiol.*, 1912, **146**, 543–552.
7. Baumann, C. Beiträge zur Physiologie des Sehens: V. Subjektive Farbenerscheinungen. *Pflüg. Arch. ges. Physiol.*, 1916, **166**, 212–216.
8. Baumann, C. Beiträge zur Physiologie des Sehens: VII. Subjektive Farbenerscheinungen. Subjektives und objektives Empfinden. *Pflüg.*

Arch. ges. Physiol., 1918, **171**, 496–499.

9. BENHAM, C. E. The artificial spectrum top. *Nature, Lond.*, 1894, **51**, 200.
10. BENHAM, C. E. The artificial spectrum top. *Nature, Lond.*, 1895, **51**, 321.
11. BIDWELL, S. On subjective colour phenomena attending sudden changes in illumination. *Proc. roy. Soc.*, 1896, **60**, 368–377.
12. BIDWELL, S. On negative after-images following brief retinal extinction. *Proc. roy. Soc.*, 1897, **61**, 268–271.
13. BIDWELL, S. Some curiosities of vision. *Notices of the Proceedings at the Meetings of the Royal Institution of Great Britain*, 1896–98, **15**, 354.
14. BILLS, M. A. The lag of visual sensation in its relation to wave-lengths and intensity of light. *Psychol. Monogr.*, 1920, **18**, No. 5.
15. BLAIR, J. M., & HYLAN, M. C. The intermittency effect in photographic exposure. *J. opt. Soc. Amer.*, 1933, **23**, 353–358.
16. BREWSTER, D. On certain affection of the retina. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1861, **21** (4th series), 20–24.
17. BROWN, C. R. A 'universal' apparatus for research in physiological optics. *J. exp. Psychol.*, 1943, **33**, 340–349.
18. BROWN, C. R. & GEBHARD, J. W. Visual field articulation in the absence of spatial stimulus gradients. *J. exp. Psychol.*, 1948, **38**, 188–200.
19. BRÜCKE, E. Über die Nutzeffect intermitterender Netzhautreizungen. *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften*, 1864, **49** (Abtheilung 2), 128–153.
20. C(ATTELL), J. McK. The spectrum top. *Science*, 1895, **2** (new series), 13.
21. CHARPENTIER, A. Phénomènes de coloration apparente observés sous l'influence d'excitations lumineuses instantanées. *C. R. Soc. Biol. Paris*, 1891, **3** (9 sèrie), 596–599.
22. CHARPENTIER, A. Remarques et expériences au sujet de la coloration entopique des lumières blanches instantées. *C. R. Soc. Biol. Paris*, 1891, **3** (9 sèrie), 601–604.
23. CHARPENTIER, A. Isolement des couleurs dans la luminière blanche par leur action successive. *C. R. Soc. Biol. Paris*, 1892, **4** (9 sèrie), 533–536.
24. COLENBRANDER, M. C. Über subjektive Farbenerscheinungen. *Acta ophthal., Kbh.*, 1933, **11**, 437–460.
25. DONISELLI, C. Sui fenomeni d'induzione cromatica da luce bianca e sulla natura dei processi consecutivi. *Arch. Fisiol.*, 1907, **4**, 561–593.
26. DOVE, H. W. Ueber Scheiben zur Darstellung subjectiver Farben. *Pogg. Ann. Physik u. Chemie*, 1848, **85** (151) 526–528.
27. DOVE, H. W. *Darstellung der Farbenlehre und Optische Studien*. Berlin: G. W. F. Müller, 1853. Pp. 281–283.
28. DUBOIS, R. Recherches expérimentales sur le rôle de la contractilité dans les mécanismes sensoriels chez les mollusques. *J. de Psychol.*, 1922, **17**, 787–805.
29. EDRIDGE-GREEN, F. W. The artificial spectrum top. *Nature, Lond.*, 1895, **51**, 321.
30. EDRIDGE-GREEN, F. W. Some subjective phenomena of vision. *J. Physiol.*, 1915–16, 50, xl.
31. ERB, M. B., & DALLENBACH, K. M. 'Subjective' colors from line-patterns. *Amer. J. Psychol.*, 1939, **52**, 227–241.
32. EXNER, S. Bemerkungen über intermittirende Netzhautreizung. *Pflüg. arch. ges. Physiol.*, 1870, **3**, 214–240.
33. FECHNER, G. T. Ueber eine Scheibe zur Erzeugung subjectiver Farben. *Pogg. Ann. Physik. u. Chemie*, 1838, **45** (121), 227–232.
34. FINNEGAN, J. M., & MOORE, B. The

artificial spectrum top. *Nature, Lond.*, 1895, **51**, 292–293.

35. Forster, G. La théorie dynamique de la vision des couleurs. *Année psychol.*, 1923, **24**, 26–69.
36. Fry, G. A. Modulation of the optic nerve-current as a basis for color-vision. *Amer. J. Psychol.*, 1933, **45**, 488–492.
37. Fry, G. A. Color phenomena from adjacent retinal areas for different temporal patterns of intermittent white light. *Amer. J. Psychol.*, 1933, **45**, 714–721.
38. Fry, G. A. Color sensations produced by intermittent white light and the three-component theory of color-vision. *Amer. J. Psychol.*, 1935, **47**, 464–469.
39. Fry, G. A. Color sensations produced by intermittent spectral stimuli. *Amer. J. Psychol.*, 1936, **48**, 326–330.
40. Gebhard, J. W. Chromatic phenomena produced by intermittent stimulation of the retina. *J. exp. Psychol.*, 1943, **33**, 387–406.
41. H., E. Expériences sur les transformations subjectives des couleurs. *La Nature*, 1897, No. 1268, 256.
42. Hannay, J. B. Colour perception. *Nature, Lond.*, 1881–82, **25**, 604–605.
43. Helmholtz, H.v. *Treatise on physiological optics.* (Vol. II.) (3rd ed.) (Ed., J. P. C. Southall). Menasha, Wis.; George Banta Publishing Co. for the Optical Society of America, 1924. Pp. 255–261.
44. Henry, C. Applications à la tachymétrie et à l'ophtalmologie d'un mode de production, jusqu'ici inexpliqué, de la couleur. *C. R. Acad. Sci., Paris*, 1896, **122**, 406–408.
45. Hurst, C. H. A spectrum top and a complementary one. *Nature, Lond.*, 1895, **51**, 510.
46. Ives, H. E. Visual diffusivity. *Philos. Magazine*, 1917, **33** (series 6), 18–33.
47. Ives, H. E. The resolution of mixed colours by differential visual diffusivity. *Philos. Magazine*, 1918, **35** (series 6), 413–421.
48. Judd, D. B. Color vision. In O. Glasser (Ed.), *Medical Physics.* Chicago: Year Book Pub., 1944. Pp. 265–282.
49. Klopsteg, P. E. A curious color phenomenon. *Science*, 1931, **73** (new series), 590.
50. Kormann, F. Die Erzeugung eines kontinuierlichen Spektrums auf Grund der Benhamschen Farben. *Z. Sinnesphysiol.*, 1931, **62**, 158–166.
51. Lemmon, V. W. Entoptic colors. *Science*, 1932, **75** (new series), 217.
52. Liveing, G. D. On Benham's artificial spectrum top. *Nature, Lond.*, 1894, **51**, 167.
53. Liveing, G. D. The artificial spectrum top. *Nature, Lond.*, 1894, **51**, 200.
54. Luckiesh, M., & Moss, F. K. A demonstrational test of vision. *Amer. J. Psychol.*, 1933, **45**, 135–139.
55. Newton & Co. The artificial spectrum top. *Nature, Lond.*, 1895, **51**, 463.
56. Pauli, R., & Wenzl, A. Über Farbenempfindungen bei intermittierendem farblosen Lichte. *Arch. ges. Psychol.*, 1924, **48**, 470–484.
57. Percival, A.S. Colour phenomena due to intermittent stimulation with light: note on the colours of Benham's top. *Trans. ophthal. Soc. U.K.*, 1909, **29**, 119–125.
58. Piéron, H. Le méchanisme d'apparition des couleurs subjectives de Fechner-Benham. *Année psychol.*, 1922, **23**, 1–49.
59. Piéron, H. Des lois du déséquilibre chromatique initial et de la prépondérance de la diffusion chromatique dans l'excitation lumineuse de la rétine. (Méchanisme de production des couleurs subjectives de Fechner-Benham). *C. R. Soc. Biol. Paris*, 1922, **86**, 922–925.
60. Piéron, H. Le mécanisme des couleurs subjectives de Fechner-

Benham. *J. de Psychol.*, 1923, **20**, 75–80.

61. Piéron, H. Influence de la composition de la luminière sur la nature des couleurs subjectives de Fechner-Benham. Données complémentaires. *Année psychol.*, 1928, **29**, 229–233.
62. Prevost, B. Sur une apparence de décomposition de la lumière blanche par le mouvement du corps qui la réfléchit. *Mémoires de la Société de Physique et d'Histoire naturelle de Genèva*, 1823–26, **3**, 121.
63. Rood, O. N. On a new theory of light, proposed by John Smith. *Amer. J. Sci. and Arts*, 1860, **30** (2nd series), 182–186.
64. R(ood), O. N. Subjective optical phenomena. *Amer. J. Sci. and Arts*, 1861, **31** (2nd series), 417.
65. Sanford, E. C. A laboratory course in physiological psychology: IV. *Amer. J. Psychol.*, 1893, **5**, 390–415.
66. Schwartze, T. *Grundgesetze der Molekularphysik.* Leipzig: J. J. Weber, 1896. Pp. 200–206.
67. Segal, J. Une nouvelle manifestation de couleurs subjectives. *C. R. Soc. Biol. Paris*, 1940, **133**, 259–261.
68. Sexton, F. P. The spectrum top. *Phys. Soc. Lond.*, 1907–08, **21**, 392–395.
69. Silberstein, L. Investigations on the intrinsic properties of the color domain. *J. opt. Soc. Amer.*, 1938, **28**, 63–85.
70. Silberstein, L. Investigations on the intrinsic properties of the color domain. II. *J. opt. Soc. Amer.*, 1943, **33**, 1–10.
71. Skinner, B. F. A paradoxical color effect. *J. gen. Psychol.*, 1932, **7**, 481–482.
72. Smith, F. J. Apparent decomposition of sunlight by intermittent reflecting surfaces. *Nature, Lond.*, 1881, **24**, 140.
73. Smith, J. On the production of colour and the theory of light. *Rep. Brit. Ass. Adv. Sci.* (Notices and Abstracts), 1859, **29**, 22–23.
74. Smith, J. On the cause of colour and the theory of light. *Amer. J. Sci. and Arts*, 1860, **29** (2nd series), 276–278.
75. Smith, N. Colour perception. *Nature, Lond.*, 1882, **26**, 30–31.
76. Stewart, G. N. Is the law of Talbot true for very rapidly intermittent light? *Proc. roy. Soc. Edinb.*, 1887–88, **15**, 441–464.
77. Stewart, G. N. Color phenomena caused by intermittent stimulation with white light. *Amer. J. Physiol.*, 1924, **69**, 337–353.
78. Troland, L. T. The enigma of color vision. *Amer. J. physiol. Opt.*, 1921, **2**, 23–48.
79. Troland, L. T. *The principles of psychophysiology* (Vol. I). New York: D. van Nostrand Co., 1930. P. 250.
80. Turner, D. The spectrum top. *Nature, Lond.*, 1895, **51**, 438.
81. Von Kries, J. In H.v. Helmholtz, *Treatise on physiological optics.* (Vol. II) (3rd ed.) (Ed. J. P. C. Southall). Menasha, Wis., George Banta Publishing Co. for the Optical Society of America, 1924. Pp. 449–450.
82. Way, E. F. Entoptic colors. *Science*, 1932, **75** (new series), 81.
83. Weyl, H. *Mind and nature.* Philadelphia: University of Pennsylvania Press, 1934, Pp. 1–19.
84. *Anon.* An "artificial spectrum top." *Nature, Lond.*, 1894, **51**, 113–114.
85. *Anon.* An artificial spectrum. *Scientific American*, 1895, **72**, 309–310.
86. *Anon.* The "artificial spectrum" or new colour top. *Engineering*, 1895, **59**, 381.

Received July 14, 1948.

AN EXTENSION OF CONTROL GROUP DESIGN

RICHARD L. SOLOMON[1]
Harvard University[2]

This paper presents an elaboration of existing experimental design in cases where an experimental group and a control group are usually employed. It is intended as a recommendation to psychologists, especially those working in these specific fields: (1) transfer of training experiments; (2) experiments on induced changes in existing attitudes, opinions, and personal values; and (3) experiments on the effects of controlled experience on responses, skills, and performances already existing in the behavior repertoire. We feel that the use of the recommended design and procedures will provide information of a valuable sort, information which cannot be obtained with the use of existing standard research designs.

BRIEF HISTORY

The history of the idea of a controlled experiment is a long one. Usually one goes back to J. S. Mill's canons (**10**) for the concept of experimental controls. Those readers interested in the various types of controls which have been suggested in the history of psychological research may be referred to Fechner (**6**), Henri (**9**), Wundt (**21**), Titchener (**17**), Baldwin (**1**), and Warren (**18**). However, if one is interested solely in the concept of *Control Group* in the history of psychology, Twentieth Century psychology contains that history. We have not been able to find a single case of the use of control group design, as we use it today, before the year 1901. Control group design seems to have awaited the development of statistical concepts which allow for the characterization of group performances in terms of measures of central tendency; and psychologists seem to have been slow to combine statistical sophistication with experimental design. We were struck with the fact that the use of an experimental group and a control group in the three types of experiments cited above is a very recent development. The earliest use of this kind of experimental design in the study of transfer of training is found in the work of Thorndike and Woodworth (**16**), published in

---

[1] The author wishes to express appreciation to Professor C. F. Mosteller and Professor S. A. Stouffer for many original and helpful contributions to this paper. Professor E. G. Boring very kindly contributed many interesting historical notes.

[2] This research was facilitated by the Laboratory of Social Relations, Harvard University.

1901. These investigators used independent educational groups as controls on other groups. They were studying the influence of various educational sequences on proficiency in various subjects. In their observations a sequence of training such as *ABC* could be evaluated against a control group with sequence *ADC*. They attributed any differences in performance in *C* in terms of differential effects of *B* and *D*. Thorndike and Woodworth were examining educational sequences as they found them. In the actual experimental manipulation of groups, in transfer of training experiments, it is interesting that control group methodology was not being used at the time that Thorndike and Woodworth published their accounts. For example, we find a typical experiment in transfer of training by Ebert and Meumann (**4**) in 1904 in which one group was used, the typical experimental group receiving this sequence; pre-test, training, post-test. This procedure was criticized by Dearborn (**3**) in 1909, and by Sleight (**14**) in 1911. That the criticism did not immediately take hold is evidenced by the fact that Ewert (**5**) in 1926 repeated the criticism, and showed why a control group was necessary. It is interesting that the first thorough transfer of training experiment using two groups seems to have been performed in 1908 by Winch (**19**). Winch used the design which is standard today; his experimental group received pre-test, training, and post-test, while the control group received the pre-test, no training, and the post-test. It is curious that Winch did not consider his design as an innovation, and in fact, did not mention it as something special when he discussed his experiment. It seems safe to assume that prior to 1900 the experiments on transfer of training were subject to Ewert's criticism. He argued that without a control group the effects of taking a pre-test are not analyzable, and that the effects of the training procedures cannot be adequately evaluated. This point has been repeated many times (see Woodworth, **20**). Examples of experiments which have been criticized on these grounds are those of Swift (**15**), Scripture, Smith, and Brown (**13**), and Ebert and Meumann (**4**), along with many others. As far as terminology goes, we find that in the middle 'twenties Ewert (**5**) and Bray (**2**) freely used the term "Control Group" for the group that received no training. Warren included this term in his Dictionary (**18**) in 1934. Yet Harriman's recent New Dictionary (**8**) did not include it. If one is still skeptical about the relative newness of the whole idea, an examination of representative psychological textbooks serves as a check. We find that the concept of the control group generally made its way into these publications between 1920 and 1930. The idea of co-twin control and litter-mate control groups entered between 1930 and 1940.

The experimental studies on changes of attitudes, opinions, and personal values have not taken much advantage of the lessons in experimental design stemming from transfer of training experiments. The common procedure has been to give a group a pre-test, using an acceptable attitude scale, then subject the group to educational procedures of some sort, and then post-test the group with the same test or an equivalent form of it. Without a control group it is impossible to evaluate either the effects of the educational procedure or the influence of taking the attitude scale. The experiment of Peterson and Thurstone (**12**) on the effects of motion pictures upon the attitudes of children represented the first serious attempt to use control groups for the purposes of evaluating the effects of the educational procedures. This was published in 1933. Murphy, Murphy, and Newcomb (**11**, pp. 946–979) in their summary of experiments on the modification of attitudes, listed many recent experiments where no control group was used.

In general, the arguments for the use of a control group apply in the same way to the three fields of endeavor listed in our Introduction. Woodworth (**20**) has included an exhaustive account of experimental design in his discussion of transfer of training experiments. More recently, Gagné, Foster, and Crowley (**7**) have reviewed this problem of experimental design in the measurement of transfer effects. Many of their conclusions apply to experiments on induced attitude changes and the acquisition of skills. To review their recent article is unnecessary; the designs for experiments are thoroughly outlined there. The important point we wish to make is that their recommendations involved the use of an experimental group which receives controlled treatment and a control group which does not. For the three types of experiment we shall discuss, the current standard designs may be tabulated as follows:

(1) *Transfer of training:*	*Experimental Group*	*Control Group*
Pre-test on Task$_1$	yes	yes
Training on Task$_2$	yes	no
Post-test on Task$_1$	yes	yes

(2) *Experimentally induced changes in attitudes, opinions, or values:*	*Experimental Group*	*Control Group*
Pre-test on Attitude$_1$	yes	yes
Education, Propaganda	yes	no
Post-test on Attitude$_1$	yes	yes

(3) *The effects of controlled experience on any existing behavior:*

	Experimental Group	*Control Group*
Pre-test on Present Behavior$_1$	yes	yes
Experimental Treatment	yes	no
Post-test on Final Behavior$_1$	yes	yes

With matched groups, or with matched tasks, both Woodworth (20) and Gagné, Foster, and Crowley (7) listed appropriate substitute designs, but all are based upon an experimental and a control group. It is understood that the two groups can be matched by one criterion or another or may be randomly drawn from the population. The pre-test and post-test do not have to be identical tests, but they can be equivalent forms based on standardization procedures, or can be equivalent in some logical manner. It is important to state that the same scale units should be involved in both the pre-test and post-test.

We have attempted to follow the history of control group methodology in a sketchy way, and have outlined what present practice seems to be. We now wish to proceed to an extension of present practice to what we believe is a more adequate design of experimentation.

AN EXTENSION OF EXPERIMENTAL DESIGN

The experimental plan outlined above possesses these possibilities for analysis: (1) assignment of a numerical value for the effects of taking the pre-test as this factor operates on the post-test scores; (2) assignment of a numerical value for the effects of the training on the magnitude of the post-test score; and (3) improvement scores based on the contribution of the training alone and the taking of the pre-test. An assumption of additivity of scores is involved here. We wish to point this out, but we feel that it is inappropriate to go into all of the ramifications of the theory of scales of measurement. The assumption of additivity allows us this privilege: the improvement score for the control group may be subtracted from the improvement score of the experimental group in order to obtain a numerical value for improvement due to *training procedure alone*. There is at present no real agreement on the permissibility of this procedure. All we can say is that the information it gives often proves useful. This procedure is discussed by Gagné, Foster and Crowley(7).

We have seen that current control group design will allow us to assign numerical values for the effects of *taking the pre-test alone* and the effects of *training alone*. The main point we wish to make is that these two items of information are still inadequate to characterize what happens in many experiments of the sorts we have discussed. We feel that the pre-test operates directly upon the effectiveness of training, or *interacts* with the training process. That is, there is a great possibility that merely taking a pre-test changes the subjects' attitudes toward the training procedures. Or it may conceivably change the "set" or attentional factors important to the effectiveness of training. Thirdly, it may actually change the manner in which the subjects "perceive" the training material. These psychological possibilities may be termed "interactions." Thus, if a group of subjects were given a pre-test on spelling ability and the test list contained a lot of *ie* and *ei* vowel pairs, it is perfectly conceivable that the subjects would "pay attention to" certain relevant portions of a spelling lesson in a manner very different than if they had not received the pre-test at all. Or, in the case of attitude experiments, the way that the pre-test items are worded might influence the manner in which the propaganda or education procedures are "received" by the subjects. We believe that this kind of pre-test-training interaction is important, and that it occurs in a large number of experiments of the variety we have listed above. We have found that these interaction effects can be quantified. The following control group design is necessary to do it:

	Experimental Group	*Control Group I*	*Control Group II*
Pre-test	yes	yes	no
Training, Education	yes	no	yes
Post-test	yes	yes	yes

The question immediately arises concerning the possibility of obtaining an improvement (or change) score for control group II. This, we believe, may be legitimately done by statistical inference, provided that: (1) the groups are either reasonably matched at the beginning of the experiment on the basis of as many criteria as possible or are randomly drawn from the population; (2) the means and variances for the experimental group and the control group I are known for the pre-test; and (3) the N's for all three groups are reliably large. This latter qualification is, of course, partly dependent upon the prior experience of the experimenter with performances of samples on the task in question. If we take these three provisions into account, we can infer what the pre-test score for control group II *would have been* had this group actually taken the pre-test. The grand mean pre-test score for the experimental group and control group I (pooled data) represents the best estimate of what the mean of the control group II would have been had this group taken the pre-test. We shall call this value "i", the "inferred pre-test score" for control group II. Table I outlines the numerical values which must be fitted into the analysis in order for the interaction effects to be quantified.

TABLE I

NUMERICAL VALUES NEEDED FOR THE ANALYSIS OF INTERACTION EFFECTS IN A THREE-GROUP DESIGN

	Experimental Group	*Control Group I*	*Control Group II*
Pre-test Means	$a_1 \pm \sigma_{ma_1}(P_1)$	$a_2 \pm \sigma_{ma_2}(P_1)$	$i=\frac{a_1+a_2}{2} \pm \sigma_{ma_{1,2}}$
Training	yes (T_1)	no	yes (T_1)
Post-test	$b_1 \pm \sigma_{mb_1}$	$b_2 \pm \sigma_{mb_2}$	$b_3 \pm \sigma_{mb_3}$
Improvement or Change in Means	$d_1=b_1-a_1$	$d_2=b_2-a_2$	$d_3=b_3-i$

Note that the inferred pre-test score, i, involves an inferred standard error of the mean obtained from the first two groups. The implications of the table above may be outlined as follows: (1) $d_3=f(T_1)$, or, the difference score for control group II is a function of the training process *alone;* (2) $d_2=f(P_1)$, or, the difference score for control group I is a function of the pre-test process *alone*, and (3) $d_1=f(T_1+P_1+I)$, or the difference score for the experimental group is a function of the effects of taking the pre-test, the effects of the training period, as well as of the interaction effects existing between the pre-test and training processes. This relationship is assumed to be additive. Here, I is the term to which we have been referring as *interaction*, or possibly, attentional, perceptual shifts due to taking the pre-test. The term I may be positive or negative, depending upon the psychological effects of the pre-test on the way the subject "approaches" the training.[3]

In summary, with reference to Table I what we have done is to assume that: (1) the improvement of control group II is due to training or education procedures alone; (2) the improvement of control group I is due to facilitating effects of the pre-test alone; and (3) the failure of the improvement of the experimental group to be equal to the sum of the improvement scores of the two control groups is a reflection of interaction effects. These effects are quantified as follows: $I = d_1 - (d_2 + d_3)$.

---

[3] There are other interactions possible, such as the training-post-test, and pre-test-post-test interactions. We shall omit such considerations here; this is equivalent to the assumption that the post-test is a measuring device of a "pure" sort, involving no learning due to its use.

There are many ways to interpret an obtained interaction. For example, if an interaction term is relatively large compared with the improvement of the experimental group, and if it is positive in sign, we might assume that the pre-test somehow facilitated the effective use of training procedures. We then might start on a search for the kinds of attitudinal and perceptual changes that the pre-test had produced, since such changes might have a generality of application for many kinds of educational procedures. In attitude testing, a negative I term would lead one to suspect that the pre-test developed some kind of resistance to the propaganda material. This problem then could be approached by new experiments. It is possible that rules for building pre-tests which make subjects receptive to education might come out of such a program. In other words, the magnitude and direction of I ought to provide some hints to the experimenter.

We have indicated the various possible psychological statuses of the interaction term, I. There still remains the challenging problem concerning the psychological status of *training alone*. It is hard to conceive of a change in behavior due to training alone. Life history has produced certain attitudes which must inevitably interact with training procedures. The problem of familiarity is important here, together with the total personality organization. In order to treat logically the concept of *training alone* it seems necessary to postulate some prior state, A, which is common to all three groups at the time of the pre-test administration. Then training may be inferred to interact with A under all conditions in the experiment. This construction enables us to state that, all other things being equal, the improvement of control group II is due to training alone or to the training-A interaction, which ever one might prefer. Parenthetically it might be added that a study of covariance during successive stages in the training procedure might reveal some of the training-A interactions. Much work needs to be done in elucidating this problem.

Rather than indulge in the listing of all of the permutations and combinations of outcomes possible within Table I let us examine a small demonstration experiment which we have completed in order to portray one use to which the three-group design may be put.

A SAMPLE EXPERIMENT, THREE-GROUP DESIGN

Two grammar school classes, a fifth and a sixth grade, were used in this experiment. The pupils in each class were assigned to the three groups required in the new design. The groups were roughly equated in spelling ability by means of teachers' judgments. The experimental group was pre-tested on a list of words of equal difficulty. Control group I was pre-tested at the same time on the same list of words. Control group II was out of the rooms at the time. Then the experimental group and control group II were given a standard spelling lesson, covering some general rules of spelling. Control Group I was out of the classrooms during this training period. Then the post-test was given to all of the groups. It was composed of the same words as in the pre-test. The results for the two grades are presented in the tables below. It is regrettable that the *N* for each group was not larger, but in order to administer this experiment properly it was necessary to restrict the sample size. If we consider the sixth grade experiment to be a replication of the fifth grade experiment, the small sample size does not loom so importantly. It is to be noted that the variabilities are not inordinately large. We are not satisfied with the sample size, but we are willing to consider this experiment to be representative, highly indicative, and capable of reproduction.[4]

The *I* term in Table II is acknowledged to be but a "best estimate." But an examination of the variabilities is enough to reassure one that the chances are very small that the population *I* is zero. Added assurance may be obtained by an examination of the sixth grade results in Table III.

TABLE II

SUMMARY OF FIFTH GRADE RESULTS OF AN EXPERIMENT IN SPELLING, USING A THREE-GROUP DESIGN

	Experimental Group ($N=10$)	*Control Group I* ($N=10$)	*Control Group II* ($N=10$)
Pre-test	yes (P_1)	yes (P_1)	no
Pre-test means	$a_1=3.2\pm0.8$	$a_2=2.8\pm0.7$	$i=3.0\pm0.6=\frac{1}{2}(a_1+a_2)$
Training	yes (T_1)	no	yes (T_1)
Post-test	yes	yes	yes
Post-test means	$b_1=9.9\pm1.6$	$b_2=3.5\pm0.8$	$b_3=11.2\pm1.2$
Improvement means	$d_1=6.7\pm0.9$	$d_2=0.7\pm0.5$	$d_3=8.2\pm ?$
Value of *I*	$d_1-(d_2+d_3)=-2.2$		

---

[4] This experiment was carried out with the aid of Mr. Richard Walk, Harvard University. We are greatly indebted to Miss Mildred March, principal of the John Ward School, Newton, Massachusetts, for her helpful cooperation, and to Miss Anne Wilson and Miss Cecilia Dunlop, teachers, for indispensable aid in carrying out the experiment.

TABLE III

Summary of Sixth Grade Results of an Experiment in Spelling, Using a Three-Group Design

	Experimental Group ($N=8$)	*Control Group I* ($N=9$)	*Control Group II* ($N=8$)
Pre-test	yes (P_1)	yes (P_1)	no
Pre-test means	$a_1=5.4\pm1.4$	$a_2=6.0\pm1.5$	$i=5.7\pm0.8=\frac{1}{2}(a_1+a_2)$
Training	yes (T_1)	no	yes (T_1)
Post-test	yes	yes	yes
Post-test means	$b_1=11.8\pm1.0$	$b_2=6.8\pm1.4$	$b_3=14.4\pm0.3$
Improvement means	$d_1=6.4\pm1.2$	$d_2=0.8\pm0.6$	$d_3=8.7\pm?$
Value of I	$d_1-(d_2+d_3)=-3.1$		

Again, we obtain a sizeable interaction term with negative sign. We have little reason to believe that this is an accidental occurrence when we consider that the fifth grade results are comparable. The interpretation of the negative interaction term offers interesting possibilities. We know that the taking of the pre-test somehow diminished the effectiveness of the training in spelling. One possibility is that the pre-test was emotionally disturbing to the children. Another possibility is that the writing down of a word erroneously on the pre-test may have produced perseverative errors. An examination of our data does not reveal this to be true for the experimental group in excess of control group I, but we have no standards against which to evaluate the occurrence of perseverative errors.

One fact stands out clearly. If we had used the ordinary two-group design (the experimental group and control group I) the effectiveness of the teaching procedures would have been erroneously underrated. The pre-test has been shown to have vitiated some of the effectiveness of the teaching method. Thus we have revealed a paradoxical case: *with the currently-used two-group design the method used to evaluate the effectiveness of training procedures actually may operate to cut down the effects of training as they are capable of being measured by ordinary procedures.* Using control group II, however, allows us to evaluate more adequately the effects of training or the success of the teaching method. The implications here for educational research seem obvious.

The Use of a Four-Group Design

Other experiments using the three-group design are in progress in the Laboratory of Social Relations, Harvard University, placing special emphasis on the problem of experimentally-induced attitude changes. Preliminary experiments have revealed that large interactions often appear in such experiments, thus demonstrating the need for control group II. In addition, the mathematical implications of the three-group method are being explored. The analysis of data presented above was the simplest possible for the purpose. It seems clear that more sophisticated statistical procedures, such as an adaptation of the analysis of covariance, may enhance the evaluation of research results. In particular, the mathematical nature of *I*, the interaction term, needs to be investigated. Its meaning is at present obscure, and the reason for this obscurity may be revealed by Table IV.

TABLE IV

SUMMARY OF MEAN SCORES AVAILABLE FOR THE ISOLATION OF INTERACTION EFFECTS BETWEEN PRE-TEST AND TRAINING IN A THREE-GROUP DESIGN

		Pre-Test	
		yes	no
Training	yes	Exper. Group	Control II
	no	Control I	??

If this table above could be filled out logically with mean improvement scores it seems clear that a straight analysis of variance, with a solution for the interaction between pre-test and training, would be feasible. But a group with no training and no pre-test is not sensible at all, with reference to the laboratory experiments with which we have been dealing. The post-test score for such a group would be in effect a pre-test, and no improvement score could emerge. Thus, that *I* is a "true" interaction magnitude may be questioned for the case of carefully controlled laboratory experimentation.

However, in the case of field studies, such as those on attitude changes in the American soldier as a result of propaganda,[5] there is an opportunity to fill in the empty cell in the four-fold table, provided that we use a fourth group, control group III. In the case of such field studies we may consider that passage of time is an important variable, since many uncontrolled events may occur. For example, we may be giving the American soldier propaganda about the great strength of our ally, Russia. Sometime between the pre-test and the post-test, Russia may lose an important battle to the Germans. This is an uncontrolled event, which we shall designate as E₁, taking place in the period of time between the pre-test and the post-test. This event is assumed to impinge upon all members of all four groups. The design of the experiment required to reveal interaction effects in such a field experiment is shown in Table V.

---

[5] We are indebted to Professor S. A. Stouffer for this information and for pointing out the necessity of a four-group design.

TABLE V

NUMERICAL VALUES NEEDED FOR THE ANALYSIS OF INTERACTION EFFECTS IN A FOUR-GROUP DESIGN

	Experimental Group	*Control Group I*	*Control Group II*	*Control Group III*
Pre-test	yes	yes	no	no
Pre-test means	$a_1 \pm \sigma_{ma_1}$	$a_2 \pm \sigma_{ma_2}$	$i=\frac{a_1+a_2}{2}$	$i=\frac{a_1+a_2}{2}$
Training	yes (T_1)	no	yes (T_1)	no
Event, E	yes (E_1)	yes (E_1)	yes (E_1)	yes (E_1)
Post-test	yes	yes	yes	yes
Post-test means	$b_1 \pm \sigma_{mb_2}$	$b_2 \pm \sigma_{mb_2}$	$b_3 \pm \sigma_{mb_3}$	$b_4 \pm \sigma_{mb_4}$
Change means	$d_1=b_1-a_1$	$d_2=b_2-a_2$	$d_3=b_3-i$	$d_4=b_4-i$
Interaction	$d_1-(d_2+d_3-d_4)$			

It is to be noted that we now infer *two* pre-test means based on statistical assumptions: those for control groups II and III. Any change in score taking place in the time between the pre-test and post-test for group III would be attributed to the effects of event E_1. Since all of the groups are exposed to this event, the following statements may be made: (1) $d_4=f(E_1)$, or, the change in attitude for group III is a function of the outside, uncontrolled event, E_1; (2) $d_3=f(E_1+T_1)$, or the change in attitude for control group II is a function of the training an the event, E_1; (3) $d_2=f(E_1+P_1)$, or the change in attitude for contro group I is a function of the effects of the pre-test and the event E_1 (4) $d_1=f(E_1+P_1+T_1+I)$, or the change in attitude for the experimenta group is due to the summative action of the pre-test, the event E_1, th training or education process, and the interaction between the pre-tes and the succeeding events in time (E_1 and T_1). Admittedly, this analysi seems destined to be cumbersome under many circumstances. But i appears necessary in order to uncover and numerically weight the interaction phenomenon. We note now that our four-fold table (Tabl IV) above can be filled out when we use this design, by inserting contr group III into the lower right-hand cell.

In view of our argument above, is it still impossible to fill in the fourth cell of the four-fold table *for the three-group design*? True, we left that table with some question marks in the fourth cell; but we believe in the light of an analysis of the four-group design, that it may be possible to do something about the missing cell. The most obvious step would be to consider the three-group design as indicative of measures *in time*. A group with no pre-test and no training could not vary with time insofar as our measures apply. We could therefore infer a zero improvement score for the fourth cell. This seems a fairly sound procedure in well-controlled laboratory experiments, but it is subject to these dangers: (1) an unknown event, E, may have affected our three groups, so that if we actually had a fourth group, a difference score would have appeared; (2) time passage, maturation, or growth processes may have made the obtained post-test score different from the "true" pre-test score. Actually, if these cases do occur, they require the four-group design in order to analyze adequately the experimental interaction. This means that many laboratory experiments in learning, motivation, nutrition, and maturation, have been inadequately designed. This is especially true of cases where some pre-condition of the animal has been measured in order to give the experimental measure a baseline. For example, in the lever-pressing experiment, we often obtain a sample of lever-pressing activity in a pre-test, to get a baseline for unrewarded behavior in this class. But these animals all grow older, and this can be called event E. Thus, in order to assess adequately the improvement or learning scores, a four-group design seems to be appropriate. We are presenting this argument for serious consideration. Actually, at present we are not certain that such elaborate control group design is absolutely necessary; but we suspect that it is, especially in cases where accurate quantification and the discovery of constants is the aim of the experimenter.

That the interaction term, I, may be of importance in evaluating research which falls within the three main areas we have considered above, seems evident. We would like to exhort psychologists who work in these fields to study further the uses of the three-group and four-group experimental designs and report the results for discussion. It seems possible that many previously unquantified effects—attentional, attitudinal, or perceptual, may be revealed and weighted by using the suggested plans. At least that is the hope we entertain on the basis of preliminary indications.[6]

---

[6] Since the time this manuscript was submitted, a noteworthy finding has been studied. We have preliminary evidence that the pre-test may operate to reduce post-test variance in studies of attitude change. This is evidence of another sort of interaction. It is, in effect, the limiting of variance by addition of restrictions of "attention."

SUMMARY

We have presented an extension or modification of currently-used control group design. We have shown that the use of a three-group design, and in some cases a four-group design, has potentialities for revealing and weighting certain interaction effects. These interaction effects may shed light on attentional, attitudinal, and perceptual factors which often are important in these three types of psychological experimentation: (1) transfer of training; (2) induced changes in attitudes, opinions, and values; and (3) the effects of controlled experience on responses, skills, or performances which already exist in the behavior repertoire. Further study of the extension in experimental design is recommended.

BIBLIOGRAPHY

1. Baldwin, J. M. *Dictionary of philosophy and psychology.* New York: Peter Smith, 1901. (See p. 229, Vol. I.)
2. Bray, C. W. Transfer of training. *J. exper. Psychol.*, 1928, **11**, 443–467.
3. Dearborn, W. F. The general effects of spaced practice on memory. *Psychol. Bull.*, 1909, **6**, 44. (Abstract.)
4. Ebert E., & Meumann, E. Über einige Grundfragen der Psychologie der Übungsphönomene im Bereiche des Gedächtnisses. *Archiv. f. ges. Psychol.*, 1904, **4**, 1–232.
5. Ewert, P. H. Bilateral transfer in mirror-drawing. *Ped. Sem.*, 1926, **33**, 235–249.
6. Fechner, G. T. *Elemente der Psychophysik.* Leipzig: Breitkopf und Härtel, 1889. (Pp. 76–87.)
7. Gagné, R. M., Foster, H., & Crowley, M. E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97–130.
8. Harriman, P. L. *The new dictionary of psychology.* New York: Philosophical Library, 1947. (See p. 87—no mention of control group.)
9. Henri, V. Revue générale sur le sens du bien de la peau. *L'année psychol.*, 1895, **2**, 295–362.
10. Mill, J. S. *Logic.* Bk. III, chap. 8, sec. 2.
11. Murphy, G., Murphy, L. B., & Newcomb, T. M. *Experimental social psychology.* New York: Harper, 1937.
12. Peterson, R. C., & Thurstone, L. L. *The effect of motion pictures on the social attitudes of high school children.* 1933.
13. Scripture, E. W., Smith, & Brown. On the education of muscular control and poser. *Studies from the Yale Psychol. Lab.*, 1894, **2**, 114–124. (Original not seen.)
14. Sleight, W. G. Memory and formal training. *Brit. J. Psychol.*, 1911, **4**, 386–457.
15. Swift, E. J. Studies in the psychology and physiology of learning. *Amer. J. Psychol.*, 1903, **19**, 201–220.
16. Thorndike, E. L., & Woodworth, R. S. The influence of improvement in one mental function upon the efficiency of other functions. *Psychol. Rev.*, 1903, **8**, 247–261, 384–395, 553–564.
17. Tichener, E. B. *Experimental psychology.* Vol. II. Quantitative experiments. Part II. Instructors Manual. New York: Macmillan, 1923. (See p. 340.)

18. WARREN, H. C. *Dictionary of psychology.* Boston: Houghton Mifflin, 1934. (See p. 60.)

19. WINCH, W. H. The transfer of improvement of memory in school-children. *Brit. J. Psychol.*, 1908, **2**, 284–293.

20. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938. (See pp. 178–181.)

21. WUNDT, W. *Grundzüge der physiologischen Psychologie.* Leipzig: W. Englemann, 1911. (See p. 399, Vol. III.)

Received October 10, 1948.

ANIMAL HYPNOSIS

THELMA T. GILMAN AND F. L. MARCUSE
Cornell Behavior Farm, Cornell University

Despite the fact that the induced immobilization of animals has scarcely been dealt with in English psychological literature and is virtually extinct as an area of experimental investigation, the problem has had a protracted German-French history. Volgyesi (**50**) claims that the first reference to immobilization was made in 1636 by Daniel Schwenter, a mathematics professor at the University of Altdorf. Most authors, however, date its inception as an investigated phenomenon at 1646, with the appearance of the "Experimentum mirabile de imaginatione gallinae Kircher" (**50**). Kircher indicated that, by holding a hen in an uncomfortable position and fixing its head, a state of entrancement could be induced when a chalk line was drawn from its beak outward. Astounded by the "miraculous" nature of the phenomenon, Kircher concluded that the state stemmed from the hen's imagination. Despite the novelty of the discovery, enthusiasm for the solution of the problem failed to find additional fertile ground, and the problem remained in obscurity.

Development in the Nineteenth and Twentieth Centuries

Not until the middle of the nineteenth century did immobilization of animals again become a source of interest and investigation. Based on Faria's (**50**) notion of the hypnotic efficacy of visual fixation, some animals were said to entrance others by looking steadily into the eyes of their prey. The literature of the period reflects the influence of Mesmer's concept of "animal magnetism." Like Mesmer and Faria, the writers perpetuated the notion that some magnetic material emanated from the magnetizer to the magnetized by means of "passes." In 1834, Wilson wrote of the effect of "passes" of a magnetic nature made upon fish, cats, dogs, hens, pigs, geese, a horse, and a lioness. Describing the effect of these "passes" on a Bantam hen and cock, he says:

> The cock at first chuckled, and made much noise; then he became quiet, and remained so; sometimes sat down and closed his eyes; but towards the end he stood upright for a considerable time, like a statue, and neither moved hand nor foot; and when the cage was taken away, he moved not in the least, allowed me to touch and pull his comb and gills, and to stroke him down without making the slightest movement of his feet, head, or neck. . . . (**52**, p. 20).

In 1854, Newman mentioned the mesmerizing of birds and mice by snakes, and the taming of horses in half an hour by means of magnetic "passes." Describing the interest in animals at this time, Newman adds:

> Animals have been frequently fascinated for purposes of experiment, and a universal rigidity of the muscles produced to such an extent as to cause them to resemble pieces of statuary, so that the animal could be taken up and its whole weight supported by one foot and this state produced and continued at pleasure (26, pp. 16–17).

In 1881, Preyer (30) had credited Braid with the observation that a phenomenon similar to human hypnosis existed in animals in moments of extreme danger. This reaction Braid called "monoideismus," a combination of cataplexy, hypnosis, fascination, and all other phenomena which temporarily disturbed the nervous system because of the limitations imposed on it by a focused attention. Braid cited the attraction of the bird to the staring snake as an example of monodeismic activity based on a single, dominant, attention-compelling stimulus. Preyer, however, emphasized cataplexy (a fear reaction) and is generally credited as proposing the idea of "fear" as a possible explanation of the condition.

With the existence of the phenomenon firmly established in many forms of animal life, the literature (38, 48) now dealt with two principal aspects of the problem: the characteristics of the trance, and the nature and function of the state itself. The predominant explanations of the function or nature of the state were: the instinct to simulate death for protective purposes, hypnotism, or the paralysis of fear (cataplexy).

In the twentieth century the work of Holmes (17, 18, 19) Hoagland (14, 15) and, to a lesser extent, Mowrer (25) and Foley (10) have helped to make a mark on the American psychological scene. Although the investigation in the United States has continued to be sparse, several studies have been reported in which an immobile state has played a secondary role. Boas and Landauer (2) note, for example, that the best readings for cardiac rate in fowls were obtained by turning the animals on their backs. They would remain in this position for 20 to 30 minutes at a time, and sometimes for as long as two hours. Recordings were then made in these inactive states. Marcuse and Moore likewise note that their tantrum pigs could be "quieted" for conditioning studies. Describing the agitation of the pig in the Pavlovian frame, the authors write:

> It was possible to quiet the animal completely merely by scratching or clipping her flank with a pair of scissors. The quieting effect was almost instantaneous. The grunting and struggling ceased, her eyes closed, her breathing became more regular and her whole body relaxed, often to the point of hanging limply in the

straps. Cessation of the clipping immediately resulted in a full-scale resurgence of "tantrum" behavior (**22**, pp. 236–237).

Stroder (**43**) notes, in a study of narcosis in frogs, that the narcotization time is shortened by turning the subject on its back. Minami and Dallenbach (**23**) made use of the immobilization technique to compare "quiet" and active periods in a study of retroactive inhibition in the cockroach.

The designation of the phenomenon. In a historical survey of the literature one finds that the name of the condition has varied considerably. It has been referred to in the earlier literature as mesmerism (**52**), and fascination or entrancement (**26**). Darwin (**8**) called it a death-feint, while others used the terms cataplexy (**30**), conscious simulation of death (**4**), and immobilization reflex (**37**). Minami and Dallenbach (**23**) called the condition akinesis, while Mangold, earlier, had called it hypnotic akinesis. The state has also been referred to as hypertonicity (**30**), tonic immobility (**10, 14, 15**), myotonia congenita (**20, 44,**) inhibitory state (**23**), or simply, inhibition (**11, 28**).

By 1938, Stroder (**43**) had written that over 100 works were concerned with the designation of the phenomenon. Steiniger (**40**), in one of the more recent papers on this problem, cites 14 such terms and attempts to distinguish between their possible meanings.

The Induction of the Trance

Common denominators or differential factors in the types of inducing stimuli have not been systematically dealt with, and one finds scattered through the literature references to a variety of methods employed by the respective experimenters. The earliest method, introduced by Kircher (**24**), consisted of holding the hen in an unusual position, with its head held down on a flat surface while a chalk line was drawn from its beak outwards. The chalk line was later shown to be an extraneous factor.

Carrington (**5**) lists three principles of induction: bodily posture, fixation of the eye, and pressure on certain nerve centers. Dearborn (**9**) considers the nature of the inducing stimulus to be tactile or visual. Hoagland (**15**) names stroking between the eyes, pressure on the thorax, or sudden dorsal position. Rabaud (**31**), on the other hand, claims that immobility is never invoked by a sensory influence, but by a sudden shock or concussion.

Foley (**10**) classified the methods into the following four principal categories:

Repetitive stimuli. The use of this method involved the stroking of body parts such as might have been used in mesmeric "passes." Mowrer (25), for example, describes stroking the comb of the rooster until it became quiet. Bonnet and Saboul (3) mention scratching the area around a frog's eyes, pinching its nostrils, or closing its eyes. Repetitive methods have sometimes consisted of "gazing" into the eyes of the animal in the fashion initiated by Braid (50), or by the sudden presentation of a light stimulus. Mowrer (25) mentioned that the sudden presentation of strong photographic lights accidentally induced immobility in his rooster, an incident which first seems to have stimulated his interest in the problem.

Another method which might be considered repetitive is the "swing" technique where the animal is generally held in the hands of the experimenter and swung gently back and forth several times (50, 51). When the animal is set down, it is usually immobilized. Parker tells of suspending a sea anemone from a hook and enclosing it in a glass cage.

> The contracted condition can be easily induced by moving the glass vessel in which the sea anemone is suspended back and forth a little, whereby the sea anemone is made to swing and strike on the sides of the vessel. By thus stimulating the sea anemone from time to time, it can be kept in a condition of continuous and extreme contraction (27, p. 51).

Marcuse and Moore (22) also made use of a repetitive technique via continuous "stroking" or scratching at the hide of a pig.

Pressure on body parts. Some investigators felt that exerting pressure on a particular part of the body, the precise part differing with the species, would induce the state (31). Minami and Dallenbach (23) induced "rest" in their cockroaches by pressing a finger lightly against the abdominal area of the inverted subject.

Inversion. Other investigators have utilized the technique of sudden inversion. Boas and Landauer (2) mention turning the hens on their backs to keep them quiet during cardiac recordings. Rode (37) turned owls on their backs to induce the immobility, while Reisinger (34) achieved the condition in this way with Indian locusts. Stroder (43) applied this method to frogs.

Restraint from movement. Although Foley cites this as only one of several methods, some degree of restraint seems to have been present in all the investigations. Marcuse and Moore (22) had their pigs restricted in a Pavlovian frame. Mowrer (25) held the rooster's legs, while stroking his comb. Reisinger (34), working with crayfish, held the head and claws motionless, or restricted movement by holding the tail. Pavlov (28) restrained his dogs by the straps of the conditioning apparatus employed. Dearborn (9) restricted his animals from movement until efforts to move ceased. Ten Cate (46) held the octopus in his hand to prevent its moving. It is not difficult to infer from the work of Bonnet and Saboul (3), that to scratch or stroke any part of the frog, the animal must have been held down, thus restricting spontaneous movement. Minami and Dallenbach (**23**), by exerting pressure on the insects, in effect, produced restriction of movement. Gilman, Marcuse and Moore (**12**) immobilized chickens by restraining movement of the head and limbs. Inverting an animal, or holding it in an unusual posture, would also produce restriction of movement. Similarly cases where "visual fixation" was thought to be effective are to be understood in terms of the physical restriction needed to permit "gazing." Steiniger (**42**) cited an example of this when he describes Volgyesi's use of the fixation method to immobilize a lion. Volgyesi sat on the back of the lion, placing his hands on both sides of the animal's head and pulling it back so their eyes could meet. Moll's reference to Kircher's original experiment gives a more detailed and informative description of the chalk-line, visual-fixation method.

> A hen is held down on the ground; the head in particular is pressed down. A chalk-line is then drawn on the ground, starting from the hen's beak. The hen will remain motionless. Kircher ascribed this to the hen's imagination; he said it imagined it was fastened, and consequently did not try to move. Mach mentions Kircher's experiment . . . as an example of the ease with which an experiment may be erroneously interpreted. For a long time the chalk-line was held to be the essential part of the experiment . . . however, it was shown that the experiment could be successfully performed by merely holding the hen down to the ground. . . . This much is certain: many animals will remain motionless in any position in which they have been held by force for some time . . . in general, the stimulus used with the animals is tactile. . . . (**24**, p. 200).

Rijlant, describing his use of Kircher's method, says:

> L'animal est imobilisé sur une table et une ligne droite est tracée à la craie devant ses yeux . . . l'animal entre en état d'hypnose et s'y maintient pendant des heures. (**35**, p. 417).

CHARACTERISTICS OF THE TRANCE

Status of the musculature. Some observations of the appearance of the phenomenon describe it as a tetanic muscle contraction (**4**, **6**, **15**, **21**, **30**), or atonicity (**1**, **45**, **46**). Most authors who have been concerned with the problem of muscle condition seem to believe that hypertonicity is characteristic of the condition. Foley says that " . . . a high degree of plastic tonus of the skeletal musculature is characteristic of tonic immobility" (**10**, p. 515). Dearborn observed that crayfish fall into a " . . . state of extreme tonicity amounting often almost to a cramp of the muscle" (**9**, p. 407). But it is also possible, as Rijlant (**35**) claims, that both a relaxed and contracted muscle condition may occur during the course of a trance. Rode (**37**) observed that the entire body of his owl contracted, and then relaxed slowly.

Heart rate. Although one finds some references to cardiac changes in the immobilized state, little of the material is quantitative. Preyer (**30**) and Verworn (**49**) observed that heart rate increased. Sidis (**39**), however, claimed his frogs showed retarded cardiac rate. Beritoff (**1**), working with several vertebrates, also noted cardiac decrease, as did Marcuse and Moore (**22**), working with pigs. Rijlant (**36**), found no change in the heart rate of an immobilized rabbit. Gilman, Marcuse and Moore (**12**) working with chickens found no significant difference in the heart rate of susceptible animals during trance, nonsusceptible, and control animals.

Respiration. Prayer (**30**) referred to noisy respiration, while Verworn (**49**) and Coriat (**6**) observed accelerated respiration. Rode wrote:

> Dès que la Chevêche est "endormie" le fait le plus frappant est la modification du rhythme respiratoire qui s'accélère, alors que la cage thoracique se dilate beaucoup plus que pendant la respiration normale (**37**, p. 452).

Parker (**27**), in his study of the sea amenome, observed that the excretion of carbon dioxide was greater in the process of contraction than in the process of relaxing; however, no significant differences in output were observed during the contracted, as compared to the relaxed, state. Sidis (**39**), Piéron (**29**) and Beritoff (**1**) claim that respiration decreased, while Ten Cate mentions that respiration was unaffected (**46**).

Analgesia. Although the majority of experimenters believe analgesia to be present in the immobilized animal, there is no unanimity of opinion. Holmes (**18**) and Darwin (**8**) both describe the absence of pain-reaction in immobilized insects, while Stroder (**43**) mentions analgesia in immobilized frogs under surgery. Ten Cate (**45**) observed surgical analgesia in angelfish, Marcuse and Moore say: "On one occasion the experimenter inadvertently cut the animal with the scissors, drawing blood, but elicited no reaction" (**22**, p. 237). Coriat, however, says that "In neither the frog nor the guinea pig were there any signs of cutaneous anesthesia during the cataleptic states" (**6**, p. 344).

Steiniger (**40**), in accordance with his concept of unimpaired sensation during immobilization, believes that pain is felt by the immobilized animal but is not overtly manifested because of its inability to move.

General reactivity. In general, animals are said to be insensitive to stimulation, although diversity of opinion exists. Holmes (**17**), describing Faria's beetle experiments, points out that the presence or absence of moving objects or sound did not affect immobility. In his own experiments with water scorpions, Holmes (**18**) found his subjects responsive to light, contact, moisture and sudden changes in temperature. Rabaud (**32**) implies decreased sensitivity when he says that contracted muscle (the immobilized state producing tonic contraction) is less excitable. Beritoff (**1**) claims there is a decrease in reactivity; however, he qualifies this by pointing out that the immobile state is composed of two phases, active and inactive, and only in the first, active phase, is decreased reactivity found. Stroder (**43**) writes that all reactions are absent during immobilization.

Bonnet and Saboul (**3**) found increased electrical stimulation necessary to elicit reactions from a frog. Foley (10) cites the "lack of overt reactivity to external stimulation" as another characteristic of the state, and, in defining the condition, calls it "negative or quiescent behavior, even in the presence of disturbing stimulation." Hoagland includes lack of reactivity to environmental stimuli as a characteristic of the state, although he adds that in lizards "the mere movement of objects in the environment was often sufficient to bring about a premature recovery" (15, p. 718). Rode (37) observed that, though the presentation of auditory or visual stimuli was ineffective, the state could be instantly terminated in the owl by touching its extremities. Reisinger (33), working with crayfish, claimed that light or darkness, and electricity were inconsequential in their effects on immobilized subjects. Bramwell (4), Verworn (49), and Steiniger (40) believe that sensation of the stimulus in an immobilized state is not diminished, even though reactivity is inhibited.[1]

Depth of trance. In relation to the study of reactive behavior the question of the depth of the trance has received scant attention. At best, Beritoff (1) merely indicates that the condition is not always at the same level from the moment of inception to the moment of cessation. His opinion, however, is without experimental support, while other investigators, apparently considering the state uniform in depth at all times, likewise show no evidence to support the contrary opinion.

Susceptibility. The problem of susceptibility as against non-susceptibility has not received very extensive attention. Both Hoagland (14) and Rabaud (32) believe that tonic immobility is of an all-or-none character. Coriat mentions that susceptibility in any one subject is variable: "Sometimes a given experiment would succeed with one animal, on other occasions the experiment would be a failure with the same animal" (6, p. 344). Darwin (8), writing of prolonged duration in some species, mentions that other varieties will not "feign" twice in succession. Holmes (17) mentioned the appearance of nonsusceptibility in scorpions at different times. McGonigal (21) mentions individual differences in susceptibility among the insect species. Dearborn observed that: "Susceptibility seems to increase distinctly with repetition of the impression on the average, but this is very erratic" (9, p. 424). He, too, notes, variation in reaction from subject to subject.

Several investigators have observed the influence of specific factors on susceptibility. Holmes (18) mentioned that induction of the trance when the subjects were under water (the natural habitat of water scorpions) was very difficult. Hoagland (**13**) claimed that amyl nitrate prohibited the appearance of the state.

---

[1] The concept of "reactivity" needs clarification. Subjects may react, in the eyes of an experimenter, because they make gross, clearly perceptible movements. On the other hand, the subjects may make movements so minimal that they are apt to go unnoticed (eye blinking, minute attempts at limb flexion, low vocalization) or else, if observed, are considered only partial movements, and thus may or may not be considered a reaction. Motor reactions, then, may be total (spontaneous rise), partial (head or limb movements) minimal, or completely lacking.

Some investigators have attempted to determine the relationship of the mechanisms of balance to the induction of the trance condition (**15**, **50**). Extirpation of the vestibular centers, however, produced no difference in the susceptibility of animals to immobilization. Still others have attempted to find a relationship between cerebral ablations and the immobility reaction. Tonkikh (**47**), working with frogs, found no change when the hemispheres were removed. Verworn, describing the tonic contraction occurring in the state, says: "Dieser charakterische Symptomen-Complex tritt auch noch ein, wenn man den Thieren das Grosshirn herausgenommen hat" (**49**, p. 154). The possible effect of taming had received no attention until the work of Steiniger (**40**). He found that birds confined to cages in a laboratory were less suceptible than birds allowed to roam freely. Comparisons of free and "tame" animals were made on the basis of the tendency to spontaneous immobility, induced by the presence of human beings. This difference, he explains, is due to diminution of fear in the "tame" animals by virtue of their relatively greater familiarity with man. Gilman, Marcuse and Moore (**12**) found that "taming," *i.e.* experience with handling by humans, significantly decreases susceptibility in chickens.

Duration. Piéron, studying frogs, said: " . . . on assure son immobilité pendant un long temps qui . . . peut devenir de plus en plus long . . . qu'on répète l'expérience" (**29**, p. 230). Moll (**24**), who also studied frogs, claimed that duration increased with repetition of the experience.

Not all writers, however, believed that duration increased with repetition. Rabaud wrote: "Il ne me parait pas prouvé que cette durée s'acroisse ni qu'elle diminue avec les répétitions" (**31**, pp. 75–76). In Holmes' earliest paper (**16**), he indicates the diminution of trance duration with repetition, and, ten years later (**18**), wrote that all the forms studied showed a gradually decreasing duration as successive "feints" were induced. McGonigal (**21**) stated that graphs of duration all showed a sharp peak at the beginning, and then declined slowly. Mowrer (**25**) also noted decreased duration. Hoagland states: "The duration of the periods of immobility vary considerably from species to species, and to a lesser degree with individuals within the species" (**15**, p. 427). He describes the character of variation of duration as varying rhythmically in most cases. So strongly did he feel the rhythm to be apparent that he explained deviations from the rhythm as predictive of death.

Other writers merely note the duration periods in their respective subjects and mention the variability of the condition from time to time in a single subject, or between individual subjects. Rode (**37**) wrote of his owl that duration was never less than 15 minutes, and often better than one hour. Volgyesi (50) also found repeated, prolonged duration in the owl. Minami and Dallenbach (23) write that the duration of inactivity in cockroaches is not long and varies considerably.

Several investigators have described the effect of particular factors on duration. Holmes (18) believed that the use of a dim light produced a longer "feint," while the trance could be shortened by the appearance of a bright light. Fabre, according to Holmes (17), found that extremes of heat or cold prolonged the trance in beetles. Reisinger (33), on the other hand, claimed that light and electricity did not affect the condition at all. Increased duration was obtained by Hoagland (13) with the administration of ergotamine and adrenalin. Rode (37) claimed that the state could be instantly terminated by touching the extremities of the owl on which he made his observations, but added that such provoked awakening was followed by a few seconds of torpor, particularly after a long immobilization. Dearborn (9) wrote that sensory stimulation precipitated the cessation of the trance. Removal of portions of the brains of water scorpions was said to decrease duration (17).

Some writers have attempted to explain the decrease in duration as a function of fatigue. Holmes said " . . . one would expect to find a diminution of the duration, and perfection of the response as a simple consequence of fatigue" (19, p. 212). Rabaud, who disputed increase or decrease with repetition, made some allowance for fatigue when he admitted: "Peut-être, cependant, intervient-il à la longue un phénomène de fatigue tel que les excitations ne parviennent plus à immobiliser l'animal" (31, p. 76). But the number of repetitions are not cited in the data, nor is one informed whether repetitions are immediately successive or extend over a period of days. Gilman, Marcuse and Moore (12) report a decrease in trance duration when immobilization trials extend over a 21 day period. When the same number of trials are given in a one day period no such decrease is found. These results are said to argue against an explanation in terms of fatigue.

Phylogenetic distribution. In Volgyesi's survey (50) of the history of "animal hypnosis," he mentions a sufficiently varied group of animal subjects to indicate the general phylogenetic spread of the phenomenon. Frolov also states " . . . the capacity of becoming immobile under the influence of compulsory restriction of movement is characteristic of all species and ages" (11, p. 158). The literature deals with the application of immobilization techniques to birds (2, 12, 41, 50, 51) reptiles such as the lizard (14), amphibians (crabs) (33), frog (54), octopus (46), insects such as water scorpions (16), cockroach (23), and a variety of mammals that includes the pig (22), monkey (10), dog (28), lion (50), and man (53).

Some writers were of the opinion that the state reached its maximum in the insect group (7, 17). Coriat felt that

> . . . the lower in the phylogenetic state was the animal experimented upon, the

> easier it became to produce a motionless state. Furthermore, the duration of the state in the crustacean was longer than in those vertebrates which had a fairly differentiated nervous system. (6, pp. 344–345).

If, however, duration of the state is the criterion, such a belief is far from tenable since extremely long periods of duration appear in many groups higher in the phylogenic scale. It would be safest to rest with the knowledge of the rather consistent appearance of the condition all along the biological scale.

It is of interest to note that an analogous condition has occasionally been observed in the human being. The condition known as cataplexy takes the form of cataleptic fits which are elicitable by sudden emotion (fear, laughter). Wilson, presenting some illustrative case histories, concludes that it may be the perpetuation of " . . . a biologically old automatism, a defense-reaction in which immobility is of advantage to the frightened animal" (**53**, p. 51).

Hoagland makes another observation on analogous behavior in humans. He writes:

> I have been able to produce the condition in adult human beings. If one bends forward from the waist through an angle of 90 degrees . . . and . . . is violently thrown backwards . . . the skeletal muscles contract vigorously and a state of pronounced immobility . . . may result (**14**, p. 716).

"EXPLANATIONS" OF THE PHENOMENON

Reflex response and cerebral inhibition. Some writers have attempted to explain the state by dealing with what was believed to be the neural mechanism producing it. Rabaud says it is ". . . une proprieté fondamentale du systeme nerveux et dérive, comme telle, de la constitution des organismes considerés; elle en est l'une des résultantes et ne saurait avoir d'autre raison d'être" (**32**, p. **141**). He goes on to explain that the alternating mobilization-immobilization impulses of the nervous system produce motion and rest, respectively, and thus the condition would arise. He sums up its biological significance by saying: ". . . l'immobilisation réflexe est un cas particuler, une simple exagération du tonus musculaire normal" (**32**, p. 128).

Bonnet and Saboul (**3**) attribute the condition to a depression of reflexes caused by the modified function of the medullary centers. Verworn described the mechanism as an irritability of the cerebrum and inhibition of the motor areas of the brain. Introducing the idea of the "Lagereflex," he wrote: "Sie ist nichts anders als die Hemmung von spontanen Bewegungen oder Handlungen . . . wenn gleichzeitig ein starker Sinneseindruck uns intensiv erregt. . . . Das Wesentliche der Erscheinung ist vielmehr der tonische Lagereflex" (**49**, p. 54).

Coriat speaks of "cerebral inhibition," but adds, "I cannot agree with the hypothesis that cerebral inhibition is a reduction of psychic intensity. This inhibition of the brain probably took place through the kinesthetic stimuli of the manipulations being poured into the central nervous system. . . . " (6, p. 346.)

Rijlant (35) believed the condition to be a result of the modified state of the central nervous system which produced a state of muscle contraction. Pavlov (28) and Frolov (11) used neural inhibition as their explanation for the phenomenon. Ten Cate (45) emphasized permanent excitation as the causative agent. Holmes (17), although using the term "death feint," rejects the explanation implicit in the term, and considers immobility to be some kind of "thigmotatic response." Verworn's concept (49) of "position-correction" is likewise reflexive.

Hoagland postulates the existence of inhibitory hormones acting on the central nervous system: "These hormones are assumed to produce tonic immobility, when present above threshold concentrations, by inhibiting impulses from 'higher nervous centers' and by allowing impulses to pass from the tonic centers to the muscles" (13, p. 842).

Death feigning. The theory of death feigning, or the simulation of death, posits the idea that the animal, in the face of danger, instinctively succumbs to immobility as a protective guise. Darwin (8) was one of the first to speak of this idea, although he notes that "fainting" with the paralyzing effects of fear may be confused with the instinctive death-feint reaction. Some doubt may have entered his thinking, for he noted, in a comparison of dead and "feigning" insects, that the immobilized insects had not adopted the characteristic death posture. Bramwell wrote:

> I am inclined to think that in many instances it is a conscious simulation of death, adopted by the animals from the instinctive knowledge of the fact that certain birds and beasts of prey, except under pressure of extreme hunger, will not attack what is dead . . . death, or catalepsy, was only shammed . . . (4, p. 156).

Criticism can be raised as to the possibility of such voluntary behavior being within the intellectual range of subhuman species.

McGonigal raises a question on the teleological value of the death feint.

> The interesting point of this whole phenomenon is that, assuming with Darwin that this process was developed as a protective instinct to frighten away enemies (a presumption not as yet superseded), in this case teleology has outdone itself, for a passionate or excitable individual often rises so vigorously and on limbs so rigid that he falls on his back. . . . (in) the power of any foe . . . comparatively helpless and exposed (21, p. 408).

This criticism would seem to be an important one, for just as one would posit an instinct to feign death, one would have to assume that animals of prey had an instinct to avoid death-feigning, or dead, animals.

Sleep. Several people have adopted the theory that the condition is sleep, or some form of it. Heubel and Wundt (**24**), Sidis (**39**) and Pavlov (**28**) support this hypothesis. Heubel and Wundt considered the resting conditions (sleep and immobility) to be the consequence of the cessation of external stimulation. Sidis, however, considered this a "hypnoidal" or sub-waking condition and not true sleep.

Hypnosis. Although the term "hypnosis" has been used by some investigators, it has customarily been used with the qualification, "so-called." Erhard (**30**) has supported this theory, but the most vigorous exponent of hypnosis is Volgyesi (**50**) who criticized sharply those who used the term, "so-called animal hypnosis." He strongly believed the condition to be the result of hypnotic influence. Bramwell, however, had earlier indicated his opposition to the hypnosis theory, writing:

> The more one appreciates the complexities of the hypnotic state—the fact that it is essentially distinguished by an increased power of controlling the organism, without diminished consciousness or volition—the more one must recognize that it finds no analogy in an animal deprived of its cerebral hemispheres, and discharging a continuous stream of nervous motor impulses from its lower centres (4, p. 157).

Spatial disorientation. Verworn (**49**), Rabaud (**31**), and Hoagland (**14**) have attempted to describe immobility as primarily related to an impairment of the customary spatial relationship of an animal and its environment.

Paralysis of fear. The principle idea presented in this hypothesis is that of a paralysis produced by fear. Preyer (**30**) first emphasized the fear factor by speaking of "cataplexy." McGonigal (**21**) implies the presence of fear, when he points out that Holmes' successful attempts at immobilizing water scorpions might not have occurred if the subjects had been permitted to remain in their natural habitat while being immobilized. Rujka (**38**) extends the fear explanation further by considering the state to be a form of anxiety reaction. Steiniger (**40**) believes the state of inhibited reaction to be attributable to two factors: the psychological factor of fear which induces the condition, and the instinctive reflexive basis for the phenomenon. Its ecological importance, he says may have two explanations. Vertebrates in this condition can, in an instance of approach or pursuit by an enemy, be overlooked by f movement on the part of the immobilized animal.[2]

Gilman, Marcuse and Moore (**12**) emphasize the importance of fear as a factor determining not only susceptibility but also duration of trance in chickens. They suggest that fear and disrupted stimulus-reception are interactive in the sense that interference with normal proprioception may evoke fearful behavior and thus both factors may be operative in eliciting the experimentally induced immobile state.

SUMMARY

It has been found that under certain conditions a "trance" state can be induced in a wide variety of animals. Attempts to picture the characteristics of this condition have, as a result of inconsistencies in technique, been discrepant in their conclusions. A variety of methods for deliberately inducing the condition have been referred to in the literature, but all seem to involve, to some degree the factor of restraint of movement. Since its description in 1646, the principal hypotheses have been death feigning, hypnosis, spatial disorientation and paralysis of fear.

BIBLIOGRAPHY

1. BERITOFF, J. Ueber die Enstehung der tierischen Hypnose. *Z. Biol.*, 1929, **89**, 77–82.
2. BOAS, E. P., & LANDAUER, W. The effect of elevated metabolism on the hearts of the frizzle fowl. *Amer. J. med. Sci.*, 1933, **185**, 654–664.
3. BONNET, V., & SABOUL, R. Contribution a l'étude de l'hypnose animale. *J. Physiol. et Pathol. generale*, 1935, **33**, 887–906.
4. BRAMWELL, J. M. *Hypnotism.* London: DeLaMare Press, 1906.
5. CARRINGTON, H. Hypnotizing animals. *Psychic. Res.*, 1931, **25**, 41–44.
6. CORIAT, I. H. The nature of sleep. *J. abnorm. Psychol.*, 1912, **6**, 329–367.
7. CROZIER, W. J. Reflex immobility and the central nervous system. *Proc. Soc. exp. Biol.*, 1923, **21**, 55–56.
8. DARWIN, C. A posthumous essay on instinct. In G. J. Romanes, *Mental evolution in animals* (pp. 360–364). New York: Appleton, 1900.
9. DEARBORN, G. V. N. Psychophysiology of the crayfish. *Amer. J. Physiol.*, 1900, **3**, 404–443.
10. FOLEY, J. P., Jr. Tonic immobility in the rhesus monkey (Macaca mulatta) induced by manipulation, immobilization and experimental inversion of the visual field. *J. comp. Psychol.*, 1938, **26**, 515–526.
11. FROLOV, Y. P. *Pavlov and his school.* London: K. Paul, Trench, Trubner, 1937.
12. GILMAN, T. T., MARCUSE, F. L., & MOORE, A. U. "Animal hypnosis": A study in the induction of tonic immobility in chickens, *J. comp. and physiol. Psychol.* (In press.)
13. HOAGLAND, H. Quantitative aspects

---

able to escape at a convenient time, it would have to be able to recognize the opportunity when it came, and be able to mobilize its muscular resources for escape. When the effects of immobilization wear off, animals frequently spring to an upright position, but there is no evidence that the immobilization disappears when the chance for flight arises.

of tonic immobility in vertebrates. *Proc. Nat. Acad. Sci., Wash.*, 1927, **13**, 838–843.

14. Hoagland, H. On the mechanism of tonic immobility in vertebrates. *J. gen. Physiol.*, 1928, **11**, 715–741.
15. Hoagland, H. The mechanism of tonic immobility, *J. gen. Psychol.*, 1928, **1**, 426–447.
16. Holmes, S. J. Death-feigning in Ranatra. *J. comp. Neurol. Psychol.*, 1906, **16**, 200–216.
17. Holmes, S. J. The instinct of feigning death. *Pop. Sci. Mon.*, 1908, **72**, 179–185.
18. Holmes, S. J. *Studies in animal behavior*. Boston: Badger, 1916.
19. Holmes, S. J. A note on tonic immobility. *J. gen. Psychol.*, 1928, **2**, 378.
20. Lush, J. L. "Nervous" goats. *J. Hered.*, 1930, **21**, 243–247.
21. McGonigal, J. P. Immobility: an inquiry into the mechanism of fear reaction. *J. abnorm. Psychol.*, 1920, **27**, 73–80.
22. Marcuse, F. L., & Moore, A. U. Tantrum behavior in the pig. *J. comp. Psychol.*, 1944, **37**, 235–241.
23. Minami, H., & Dallenbach, K. M. The effect of activity upon learning and retention in the cockroach. *Amer. J. Psychol.*, 1946, **59**, 1–58.
24. Moll, A. *Hypnotism*. London: Scott, 1910.
25. Mowrer, O. H. A note on the effect of repeated hypnotic stimulation, *J. abnorm. soc. Psychol.*, 1932, **27**, 60–62.
26. Newman, J. B. *Fascination or the philosophy of charming*. New York: Fowler Well, 1854.
27. Parker, G. H. The excretion of carbon dioxide by relaxed and contracted sea anemones. *J. gen. Physiol.*, 1922, **5**, 46–64.
28. Pavlov, I. P., & Petrova, M. K. A contribution to the physiology of the hypnotic state of dogs. *Character and Pers.*, 1934, **2**, 189–200.
29. Piéron, H. Le problème physiologique du sommeil. *Libraries de l'Académie de Médecine*, 1913, 230–235.
30. Preyer, W. *Die Entdeckung des Hypnotismus*. Berlin: Gebrüder Paetel, 1881.
31. Rabaud, E. Le phénomène de la "simulation de la mort," *C. R. Soc. Biol. Paris*, 1916, **79**, 74–77.
32. Rabaud, E. L'immobilisation réflexe et l'activité normale des arthropodes. *Bull. Biol.*, 1919, **53**, 1–149.
33. Reisinger, L. Hypnose des Flusskrebses. *Biol. Zbl.*, 1927, **47**, 722–726.
34. Reisinger, L. Katalepsie der indischen Stabheuschrecke, *Biol. Zbl.*, 1928, **48**, 162–167.
35. Rijlant, P. Étude chez la poule des activités toniques et contractile du muscle strié pendant l'hypnose. *C. R. Soc. Biol. Paris*, 1933, **113**, 417–421.
36. Rijlant, P. Le tonus musculaire chez un mammifère en état de l'hypnose. *C. R. Soc. Biol. Paris*, 1933, **113**, 421–424.
37. Rode, R. Observations sur le phénomène d'immobilisation réflexe chez la chouette chevêche, Athene noctua Scop. *Bull. Soc. Zool. France*, 1930, **55**, 451–454.
38. Rujka, T. Beiträge zu den biologischen Grundlagen des Zusammenhanges zwischen reflektorischer immobilization und Angstzuständen. *Riv. Biol.*, 1938, **26**, 317–342.
39. Sidis, B. An experimental study of sleep. *J. abnorm. Psychol.*, 1908, **3**, 1–32.
40. Steiniger, F. Die Biologie der sogenannte "tierischen Hypnose." *Ergebn. Biol.*, 1936, **13**, 348–451.
41. Steiniger, F. Der Eingluss der Zahmheit auf die sogennate "tierischen Hypnose" bei Vögeln. *Z. Tierpsychol.*, 1941, **4**, 260–271.

42. STEINIGER, F. Eine Erwiderung auf das Buch von Franz Volgyesi, Menschen- und Tierhypnose. *Z. Tierpsychol.*, 1941, **4**, 272–280.

43. STRODER, J. Ueber der Einfluss der "tierischen Hypnose" auf den Ablauf der Narkose am Kaltblüter. *Schmerz Narkose-Anaesth.*, 1938, **11**, 82–84.

44. SUTHERLAND, G. F., & CURTIS, Q. F. Myotonia Congenita in the goat. *Proc. Soc. exp. Biol., N. Y.*, 1938, **38**, 460.

45. TEN CATE, J. Sur la production de ce qu'on appelle l'état d'hypnose animale chez la raie. *Arch. néerl. Physiol.*, 1928, **12**, 188–190.

46. TEN CATE, J. Nouvelles observations sur l'hypnose dite animale. État d'hypnose chez octupus vulgaris. *Arch. néerl. Physiol.*, 1928, **13**, 402–406.

47. TONKIKH, A. V. The role of the autonomic nervous system in the phenomenon of the so-called "animal hypnosis." *Fiziol. Zh. S.S.S.R.*, 1938, **24**, 367–371.

48. VERGER, M. P., & LAFON, M. J. La cataplexie. *Encéphale*, 1939, **34**, 121–148.

49. VERWORN, M. Die sogenannte Hypnose der Thiere. *J. Physiol.*, 1899, **23**, 53–54.

50. VOLGYESI, F. *Menschen- und Tierhypnose.* Leipzig: Fussli, 1938.

51. WARNKE, G. Zur Akinese bei jungen Möwen und Störchen. *Zool. Anz.*, 1937, **118**, 17.

52. WILSON, J. *Trials of animal magnetism on the brute creation.* London: Sherwood, Gilbert and Piper, 1839.

53. WILSON, S. A. K. Cataplexy. *J. Neurol. Psychopath.*, 1933, **14**, 45–51.

54. YERKES, R. M. The instincts, habits and reactions of the frog. *Psychol. Rev.*, 1903, **1**, 579–597.

Received November 14, 1948.

BOOK REVIEWS

EVANS, R. M. *Introduction to color.* New York: Wiley, 1948. Pp. x+ 327.

Once in a blue moon a book appears which brings together many important developments in a field at a time when integration is imperatively needed. The subject of color embraces a vast domain, ranging from physics to psychology on the scientific side and from industry to art on the applied side. The author of this book, who is Head of the Color Control Department of Eastman Kodak Company, is well equipped for his task with a background of training in optics and photography, with industrial experience in still and motion picture photography, and with an active interest in what the psychologist has to offer in unravelling the riddles of color. As a result this book deals not only with basic physical considerations but also includes a good deal of phenomenological material. Applications to color photography, art, and design are not overlooked. Two of the most difficult aspects of the subject for the lay reader, mathematics and a highly technical nomenclature, are adroitly side-stepped without any sacrifice in clarity and accuracy of exposition. All in all this is a notable contribution to visual science.

It is not possible within the confines of this review to give an adequate picture of the careful, detailed discussion of fundamentals and the vast amount of information this work contains. Anyone working seriously in the field of color today must understand how visual stimuli are measured and specified, either in terms of the I.C.I. standard observer or by means of an equivalent system; he must be aware of and on guard against errors inherent in photometric, spectrophotometric, and colorimetric measurements, especially with respect to the newer fluorescent materials coming into use; and he must have some knowledge of the vagaries of light, paint, and pigment mixtures. Certain phenomena, like dichroism and metamerism, if not properly understood, can be the source of faulty inferences regarding both the nature of the stimulus and the behavior of the eye. These and many other matters concerning light sources and reflecting, transmitting, and mixture characteristics of solid, liquid, and colorant substances are fully discussed and illustrated by means of excellent diagrams and superb colored plates.

The power of a quantitative treatment of color comes out particularly in the many uses to which the I.C.I. mixture diagram can be put. Not only does it provide a means of specifying dominant wave-length and purity of stimuli and the results of mixture of any number of lights, but it also shows changes in chromaticness of objects in different illuminants and of mixtures of dyes, paints, or pigments in various amounts or concentrations. Maximum purity theoretically obtainable from surfaces of any reflectance are thereby easily and accurately diagrammed. In spite of its limitations this method of specification of stimuli should be more widely known and used in psychology.

Along with detailed handling of the physical conditions under which colors are seen will also be found a wealth of phenomenological observations intimately related to physical and physiological antecedents. The discussion is not limited to color in the narrower sense, for matters like the perception of depth, the texture of surfaces, and the nature of shadows are explained by reference to position and size of light source, the reflecting properties of objects, and the like. It is in this area, the meeting place of psychology and physics, that this reviewer believes psychologists will find the book of greatest interest and value even though the author's concept of psychology does not seem wholly satisfactory. Let us turn to this question since it is basic to most of the work being done in the field of color whether by psychologists or others.

The key to Evans' treatment of color as most psychologists understand the concept is found in the position adopted in the 1943–44 Report of the Colorimetry Committee of the Optical Society of America wherein the subject is divided into three parts: (1) physics, which deals with sources and modifiers of light energy; (2) psychophysics, which deals with the evaluation of radiant energy by the eye under a single, specified set of conditions, viz., matching small foveal stimuli against a completely dark background; and (3) psychology, which deals with factors like the modes of appearance of Katz, so-called perceptual attributes of vision like texture, size, location, flicker, lustre and glossiness, and other factors like attention, feelings, attitudes, and intentions as they affect colors. Psychology is regarded as having to do only with subjective, qualitative matters. This position will come as a shock to those who believe in and work in psychology as an experimental, quantitative science.

At first sight the three-fold division of color seems as if it might offer a good way out of many of the difficulties which beset workers in the field of visual science. But as the realm of physics is left behind and the organism enters more and more into the total process, the split between psychophysics and psychology seems arbitrary and confusing. The mental keeps popping into the exposition, upsetting the scientific applecart again and again. At times the author himself is not sure of the validity of his fundamental position, *e.g.*, when he says: "This difference (in the way of looking at the stimulus) suggests that what has been described as a difference in attitude may, in the last analysis, be found to be a difference primarily in the way the eyes are used rather than in the mental phase. A consequent decision that the effect is purely psychophysical may be involved" (p. 166). To the psychologist it is obvious that the definition of psychophysics adopted by the Colorimetry Committee is too narrow and leads to insuperable difficulties for any consideration of color based on it. Since the time of Fechner a number of psychophysical methods other than the null method of the standard observer have been found to yield statistically reliable results in studies of vision, hearing, kinaesthesis, taste, and smell. That some of the difficult problems of colorimetry, *e.g.*, the problem of color differences, and so-called color appearances, will not be solved by taking results from the method of just noticeable differences or the null method of gauging the spectrum, and forcing them to yield information about equal sense distances or equal ratios, must be clear to all by now. A definition of psychophysics that includes the procedures made possible by the work of such psychologists as Urban, S. S. Stevens, Guilford, and certain contributions of Thurstone and others might prove very useful in approaching color problems now relegated to the psyche.

On the other hand, due weight is given in this book to factors like adaptation level and contrast in explaining color phenomena which many psychologists have been content to label "constancy" and let go at that. Part of the reason for abandoning these concepts when dealing with phenomena of constancy and color conversion may be due to the fact they were conceptualized too simply. As Evans points out, there are at least three types of adaptation: general, local, and lateral (with temporal effects to complicate each, it may be added); and three types of contrast: subject, object, and illumination contrast. Whether or not these types of adaptation and contrast are merely descriptive or are fundamentally different is still an open question.

Many psychologists will miss consideration of color theories and a good deal of information which physiology supplies in connection with this topic. Evans seems to find the tri-receptor theory plus a brightness receptor and a new concept, "cross talk" among nerve fibers, adequate for explaining a number of crucial phenomena. Until an alternative theory to the Young-Helmholtz theory comes along which is equally simple conceptually and as convenient for quantitative work we may expect people to go on using the older formulation in spite of its known shortcomings.

To sum up: this book constitutes a very important contribution to visual science and will exert considerable influence for a number of reasons. First, it makes available a great deal of material scattered in scientific, technical, and trade publications. Second, it relates a large number of phenomenological data to their antecedent physical conditions. Third, it opens up many possibilities for further investigation to the discerning reader. That the psychologist can play an important role in the future of visual science no one can doubt after reading this book, because it shows that many of the problems are predominantly psychological in character. Progress will come through the application of quantitative methods to color phenomena now relegated to mere "appearance," or to subjective factors like intentions or illusions. Evans has brought visual science farther along the way by showing that a good deal that used to be thought to be "purely psychological" is amenable to objective treatment. Much still remains to be done, however. If we may regard this book as a contribution by a physicist with a strong psychological slant, then we may hazard the opinion that there is room for another book by a psychologist with a strong physical slant!

HARRY HELSON.

Bryn Mawr College.

DOOB, LEONARD W. *Public opinion and propaganda.* New York: Holt, 1948, $3.75.

Teachers of courses in the psychology of public opinion and propaganda surely need a good textbook. If Doob's new text does not fulfill this need completely, it is nonetheless—and by far—the best text to be had.

A word first about its point of view and organization. The author takes as his systematic theme, announced in the opening chapters, the behavior theory of Clark Hull and the Yale School. With varying success, the chapters which follow attempt to apply Hullian principles to an understanding of attitude formation and maintenance, the development of public opinion, the principles of propaganda, and the myriad of behaviors which constitute mass communication. There are twenty-one chapters covering such topics as the cultural background of public opinion, sampling of opinion, the techniques of polling and intensive surveying, the nature of propaganda, content analysis, the perception of propaganda, the relation of learning and personality to propaganda reaction, and a triad on press, radio, and "the media of sight and sound."

Evaluation of Doob's book must proceed on two levels. In the first place, it is more than a textbook, representing as it does one of the few systematic attempts to apply a consistent theory of behavior to the vast realms of opinion and communication. And then, it is also a textbook.

Consider Doob's book first as an attempt at systematization. In the opening chapter, Doob urges that, in the study of opinion and propaganda, we deal not with aggregate entities, but rather with the principles of individual behavior. Let us, he says, talk not of culture or society or social pressures in the abstract, but in terms of their impact on individual behavior. The necessary equipment for such an analysis of individual behavior, Doob feels, can be stated in terms of these basic principles: the S-R relationship; the concept of drive as a tension state; the law of effect stated in terms of tension-reduction; the intervening variable of personality which comprises the organization of tendencies, which mediate between S and R; the concept "attitude," of which more, shortly; the Hullian doctrine of habit and habit organization; and finally knowledge, variously denoted as know-how or skills or subjective, cognitive context. Attitude, obviously the central term, is defined in simple terms as "the socially significant internal response that people habitually make to stimuli." "Public opinion refers to people's attitudes on an issue when they are members of the same social group." That, in essence, is the conceptual core of Doob's approach, and to it he sticks with striking tenacity through chapters ranging all the way from the learning of propaganda to the printed media.

In Chapter 5, a crucial one dealing with the "behavior of public opinion," we see a rather crucial test of the adequacy of 190 proof Hull as applied to public opinion. The result is curious. One can go only so far, in the kind of theorizing which accounts for the development of internal responses (attitudes) in terms of learned drive-reducing behaviors, of successful propaganda as drive-reducing stimuli, and so on. When Doob comes down to an attempt to state some laws about public opinion, however, we discover that he too is strapped, for, lo and behold, there appears before us a description of opinion formation mechanisms in terms of the classical mechanisms of ego-defense: rationalization, displacement, compensation, projection, etc. The chapter concludes with some stunningly thought-provoking general laws of public opinion which attempt to relate the formation and expression of opinion to such phenomena as frustration, anxiety, and the need for conformity. For the reviewer, the chapter appeared as the theoretical high point of the book, the one point at which he felt liberated from the interminable circle of things working because a habit existed, because a stimulus had reinforcing value, or because a bit of behavior brought about a reduction in drive state directly or indirectly. The theoretical model, rigorous though it may be, and useful though it has proved in certain restricted forms of behavior analysis, is a woefully impoverished one when placed alongside the real thing of public opinion and propaganda. While there are some conceptually provocative sections, one does not feel that Doob has succeeded in doing theoretical justice to the richness of the material which he treats on the descriptive level. For at this level, Leonard Doob is a knowledgeable, shrewd, and amusing observer who sees and understands much.

We must talk now about Doob's work as a textbook. It is a good textbook, and it is packed with a great deal of recent material on problems in opinion and communication. A review of this sort cannot cover all of this material and its organization in the book. Let us merely try to touch on a few of the high points, some low points, and then try to draw up a balance. The chapters on the sampling of opinion and the mechanics of polling are excellent, better than anything available for large class use. They are informative, full, and direct in exposition. So too, the chapter on the importance of public opinion, which represents a thoughtful departure from the arrant romantic nonsense which makes of public opinion an infallible demigod. The sections on propaganda have an admirable internal organization which takes the reader through propaganda from its perception, through the role of personality in mediating response, through the learning of response to propaganda, to propaganda-induced action. The point of view represents, however, little change from Doob's earlier book on propaganda. Its organization, however, is much tighter.

Now the low points. The chapters on the printed media and on radio are more impoverished than they need be. There is insufficient attention to studies of readership of newspapers, and the treatment of book reading has no reference whatever to the major studies in the field: the work of Waples and the national survey of Link and Hopf. The chapter on radio is good on statistical description of listening, but fails to take advantage of the rich literature dealing with gratification studies, the "why" of radio listening. In place of reviewing these, Doob gives a rather spotty presentation of his own propaganda principles applied at random.

But the biggest failure of the book as a textbook is its failure to present adequately, and in context, the major studies which have made of public opinion research something more than a collation of newspaper clippings from the Gallup Poll. The chapter entitled "Evaluation of Polling" is an excellent critique of some of the adolescent extravaganzas of the pollsters—in which, unfortunately, various people are pilloried for views they never thought they had—but this good criticism is not counterbalanced by the presentation of more basic and constructive work. Why not some extended treatment of such monographic studies as Lazarsfeld, Berelson, and Gaudet's *The People's Choice;* of Jones' *Life, Liberty, and Property;* of Merton's *Mass Persuasion;* of Murray and Morgan's *Clinical Study of Sentiments;* of Newcomb's *Personality and the Social Pattern;* of Likert's Federal Reserve studies; of Link and Hopf's *People and Books*? And why so little on attitude scaling? Doob seems to preoccupied at times with his conceptual task that he has tended to forget the task of presenting to his student readers a full picture of the substantive content of the field of public opinion and propaganda.

The balance, unfortunately, is not as high on the plus side as one might have expected in a work by so talented an analyst as Doob. We still lack an ideal textbook. But, for all its shortcomings, Doob's *Public Opinion and Propaganda* is clearly above the level of previous textbook attempts in this field.

JEROME S. BRUNER.

Harvard University.

SHERIF, M. *An outline of social psychology.* New York: Harper, 1948. Pp. xv+479.

This text is intended as an introduction to social psychology and presents Sherif's approach to the subject as developed in his research and as expressed in his earlier writings, in particular, *A Study of Some Social Factors in Perception*, *The Psychology of Social Norms*, and (with H. Cantril) *The Psychology of Ego-Involvements*. The book is evidently meant for students who have had a course in general psychology, since it assumes considerable knowledge of psychological terminology and theory.

Sherif defines social psychology as dealing with "the experience and behavior of the individual in relation to social stimulus situations" (p. 1), namely "interpersonal relationships, group interactions and their products." These products include "values or norms, language, art forms, institutions and technology." For Sherif, the impact of social stimulus situations on the individual constitutes the process of socialization—"the central theme of social psychology."

Throughout the text, attention is paid to two chief determiners of the individual's perceptions, thoughts and behavior: his *motives* and his *reference groups*. The *norms* of his reference groups, past and present, have largely determined his *frames of reference* comprising his *ego-attitudes* (or *values*). However, at any given time, his *deprivations* (of *biogenic* and *sociogenic motives*) are critical selective factors in his perceptions, thoughts and behavior.

Sherif holds to the view that social psychology uses the same concepts as general psychology but "extends" them to the social field. Accordingly, he most frequently introduces topics by reviewing relevant analyses and research findings of general psychology, then relates the concepts thus explored to social contexts.

In the main, social psychologists have concerned themselves with three major themes: *personality structure and functioning* (apart from personality development), *socialization* (i.e., personality development) and *group dynamics* (the interaction of individuals in face-to-face groups). Sherif places almost exclusive emphasis on socialization, largely ignoring the other two themes. Nowhere does he venture a unified account of personality structure, though he uses such structural concepts as motive, perception, attitude, ego and ego attitude. He approaches a unifying concept most closely when he defines the ego as a "constellation of attitudes in the psychological make-up of the *human* adult" (p. 252). The nature of this constellation he does not specify. Likewise, Sherif makes no attempt to describe or explain how face-to-face groups are formed, or how individuals interact in such groups. This is not to say that he avoids mentioning such groups. Indeed, he talks about crowds, in-groups and out-groups, etc., and reports some of the research studies involving different sorts of groups.

The point is that he does not go beyond a very unsystematic treatment of group structures and processes.

In his discussion of socialization, Sherif emphasizes two things: the part played by individual factors such as food deprivation in determining social perceptions and attitudes, and the effects of group norms on individuals. His account of socialization is more descriptive than explanatory. In the main, he contents himself with showing that ego involvement occurs through social experience, and that group norms and individual ego attitudes are correlated. He brushes aside such approaches to explaining personality development as learning theory and Freudian theory. While he obviously inclines toward Gestalt "field theory," he does not discuss it systematically.

Sherif cites quite a number of research studies at length. He includes detailed accounts of the University of Minnesota starvation experiments, Newcomb's Bennington study (a section written by Newcomb himself), W. F. Whyte's study of street-corner gangs, and the leadership studies of Lewin and his associates. He devotes a good deal of space to his own experiments on autokinesis and ego-involvement.

The volume is quite uneven both in style and level of difficulty. Some of the passages seem to be directed toward the professional psychologist rather than the undergraduate student. Many of the concepts used throughout the book are not carefully defined. Sherif quotes from authorities (notably C. T. Morgan, A. Gesell and E. C. Tolman) more frequently than seems necessary in his discussion of the biological bases of social behavior. On the favorable side, Sherif makes effective use of novel and highly interesting illustrative material, much of it in the form of anecdotes relating experiences of participants in World War II.

There is virtually nothing in this book on the methods of social psychology. Even while describing experiments in detail, it is the results, not the methods, which Sherif stresses.

Certainly this is not the basic textbook in social psychology which we have all been waiting for, or trying to write. However, Sherif's approach to the subject is original and provocative, and his research contributions substantial. Many instructors who admire Sherif's thinking and research on frames of reference will welcome this volume since it makes his views more accessible to the undergraduate student than his earlier writings.

GLEN HEATHERS.

New York City.

HALSTEAD, WARD C. *Brain and intelligence: A quantitative study of the frontal lobes*. Chicago: Univ. Chicago Press, 1947. Pp. 206. $6.00.

Within the so-called modern era of scientific inquiry there have been many attempts to find physical and physiological correlates of intellectual functions. On the physical side, these range from the pseudoscientific approaches of Gall and Spurzheim, early in the nineteenth century, to the more refined anthropometric studies of recent date, some of which have culminated in general typologies. Among the physiological correlates sought may be mentioned reaction time, reflex time, chronaxie, metabolism, brain waves and others. None, as a separate measure, has proved effective or convincing. Yet, somewhere within the complex matrix of the living organism must lie the substratum upon which intelligent behavior depends for its development. In phylogenetic retrospect, the highest level of the central nervous system has seemed to many the most likely place to look because of its progressive hierarchy of integrative functions.

This is precisely what Halstead has done. Exercising courage, conviction and patience he has during the past ten years or so laid the ground-work for an entirely new conception of intelligence. In *Brain and Intelligence* he tells, albeit somewhat dramatically, how he has been able to utilize neurosurgical and brain-injured patients to tease out factors in a "biological intelligence" by some unique and some traditional test procedures. Halstead is a pioneer in a very important field of endeavor and he has made substantial progress in opening up this area of investigation. He has coupled quantitative, experimental methods derived from psychology and physiology with a judicious clinico-pathological insight developed through a period of concentrated work with neurological and psychiatric patients.

In view of Halstead's extensive and significant work in forging a pathway in a virtual wilderness, this book is somewhat of a disappointment. It is stimulating and provocative, but it is disappointing because it does not make clear the nature and significance of the basic concept which underlies the work, namely that a "biological intelligence" exists. Although it is evident that a great deal of careful and ingenious experimenting has been done, the book is clearly deficient in the portrayal of the experimental procedures and in the presentation of the experimental results. On the other hand, confusing hypothetical issues are given undue space. For example, there is a persistent effort to make the facts fit certain psychoanalytic concepts; this effort is obviously strained and without proportionate success.

Part I of the book is concerned with the structure of biological intelligence. It begins with an ambition to define and measure biological intelligence. Although the desire to be clear and straightforward about biological intelligence still persists in Chapter VI with the statement, "our objective in this chapter will be a definition of biological intelligence in operational terms," there is nowhere in the book, a clear statement, operational or otherwise, as to what is meant by it. The first four chapters, or nearly one fifth of the textual part of the book, are required to identify Freud's ego with biological intelligence and to dispose lightly of previous theories of intelligence, psychometric, clinical, and neurological.

Chapter V briefly discusses the background for the present study and the method of selecting control subjects. The principle of "blended controls" was settled upon. These consisted of 14 "normals," 10 military patients with mild psychoneurosis, and 6 miscellaneous subjects under some psychological or biological stress at time of examination. One may question this procedure, as well as the matter of basing the factor analysis upon 50 military head injury cases, although reportedly recovered from mild concussions.

There is needless confusion throughout the book in reference to the number of subjects and tests involved. A total of 237 patients and control subjects and 27 tests (or indicators as Halstead prefers to call them) are repeatedly referred to, but Appendix D presents raw data for only 30 "controls" and 50 neurosurgical patients based upon 10 tests. Likewise, the summary implies that the results are based upon 237 subjects and 27 tests.

As an example of the sparsity of experimental data, Chapter VI presents the quantitative results and is composed of four and one-half pages and two tables. One is a table of intercorrelations for 13 tests or indicators, and the other contains the factor loadings for these tests as determined by the Thurstone method. Two other factor analyses supplied by Holzinger are not presented. Although these reportedly gave four-factor descriptions, the reader is not told how they agree with the four-factor solutions by the Thurstone method.

Chapters VII to X are devoted to a description and explanation of the four factors, whereas Chapter XI fits them into the nuclear structure of the ego. The four factors are: (1) a central integrative field factor (C), described as the organized experience of the individual; (2) an abstraction factor (A) found in grouping to categories and in understanding similarities and differences; (3) a power factor (P), a dynamic, energizing factor, said to counterbalance or regulate affective forces and permit further ego differentiation; and (4) a directional factor (D), which specifies the sensory modality of experience and the motor outlets—an exteriorization of biological intelligence.

Part II deals with the representation of the basic factors in the brain, the way these factors may be modified by experimental anoxia, and the differential impairment of the factors with different brain lesions. The latter aspect has been quantified in the form of an *impairment index*. The impairment index consists of a ten-point scale, one point being allotted for reaching a set criterion on each test. The score range is 0.0 to 1.0. Control subjects have a low index, whereas patients with frontal lobectomies have the highest indices. Frontal lobectomies show a higher impairment index than do equivalent ablations in other areas of the brain. It should be pointed out that only eight of the original 13 tests used in the factor analysis appear in the impairment index. Unfortunately, several errors are to be found in the impairment indices of tables 15 and 16. Also in Table 16 there are several instances of the calculation of an impairment index based on only three, four, or five of the test indicators out of a total of ten. This does not seem justified, particularly when different subjects might thus be compared on the basis of entirely different indicators. Appendices A and B provide, respectively, a brief history and diagrams of brain lesions for each of the 50 lobectomies.

Although the reader will find much to stimulate his thinking in this book, he will undoubtedly yearn for more adequate descriptions of the indicators and procedures for using them, for more data than are provided in the limited tables, and for more conclusive proof of the four factors and their significance and meaning. In general, he will probably wish that Halstead had described his tests and results completely, and perhaps had left ramifying theory and arm-chair speculation to be covered in a separate volume so labelled.

DONALD B. LINDSLEY.

Northwestern University.

Vol. 46, No. 3 May, 1949

Psychological Bulletin

PROBLEMS OF PSYCHOMETRIC SCATTER ANALYSIS

JOSEPH JASTAK
Delaware State Hospital

INTRODUCTION

Whether tests are administered in omnibus age scales like the Stanford-Binet or in homogeneous point scales like the Wechsler-Bellevue, their successes and failures always exhibit a measurable range of difficulty and a characteristic scattering of responses. The scatter records of individual cases depend partly on the way in which the scale is standardized and partly on the personality problems of the examined person. The possible clinical value of scatter analysis has been hinted at by the earliest experimenters in psychometrics. Its qualitative evaluation has gained momentum in recent years, reaching its culmination point in Rapaport's publication (26) on diagnostic psychological testing and Thurstone's volume (33) on factor analysis.

The main reason for the psychologist's persistent preoccupation with test scatter is the amount of valuable information it yields as a supplement to quantitative indices of brightness. Personality diagnoses often have a basis in the so-called intuitive interpretation of scatter whether clinicians are aware of it or not.

THEORETICAL CONSIDERATIONS

The clinical usefulness of scatter is in part contingent upon the theoretical viewpoint held by the examiner. On the other hand, the systematic study of scatter may significantly contribute to future changes in psychometric theory (19).

Two general approaches to the problem of response variations on tests are distinguishable. Brody (6), Roe and Shakow (27), Richmond and Kendig (17) and those who standardize tests (32, 34) cling to the idea that psychometric tests measure intelligence. Mental disorders reveal themselves, according to these students of scatter, through their indirect effects on cognitive processes measured by tests. Those engaged in factor analysis (33, 37) hold similar views in the interpretation of group factors. They consider the extracted factors to be organizational features within the sphere of intelligence. Though non-intellectual personality traits and group factors are being mentioned together with increasing frequency, their relationship is rarely clarified to the satisfaction of the applied psychologist.

Bijou (5), Jastak (13, 14, 15, 16), Piotrowski (22), Rapaport (26), and Schafer (29) either imply or express the view that intrinsic personality traits may be appraised through test scatter directly without the medium of intellect. Speaking of signs of invalidity, Piotrowski (22) finds that some test failures may not be regarded as legitimate symptoms of lack of intelligence. Instead, they are expressions of unorganized or disordered personality functioning in the cognitive, affective, and instinctual areas of behavior. Rapaport (26) and Schafer (29) point out the dynamic character differences reflected in the relationships between scores of various homogeneous scales. Jastak (16) finds that intelligence accounts for only 20 to 25 per cent of the variance of any one test and that the remaining variance must be accounted for by attributes and factors independent of intellectual level.

The differences between these two schools of thought may not appear to be great. Members of both groups accept scatter analysis as a major clinical tool for the understanding of individual differences. However, the personalistic approach probably assumes less and explains more. Every psychometric response has some sort of clinical validity. What it is valid for may not be accurately determined by a priori assumptions but by actual analysis. The traditional pars-pro-toto theory of intelligence artificially limits the experimental horizons of applied psychology. It ignores the Gestalt principle that every human act is the result of the total personality complex. If such traits as introversion, aggression, compulsion, suggestibility and instability affect human adjustments, it is unthinkable that they do not have a direct influence on test scores.

The intelligence theory often falsely identifies cognition with intelligence, confusing an overt act of behavior with a scientific abstraction. It also leads to the juxtaposition of intelligence and emotion which again fails to heed the distinction between theoretical constructs and concrete mechanisms (8). Furthermore, it delays the experimental verification of the theory of functional unities or independent traits which are likely to be productive of more fertile ideas in psychometrics than has been the rigid and monodimensional intelligence hypothesis.

Let us assume, for the sake of argument, that behavior is determined by two traits: (1) degree of intelligence, and (2) degree of sanity. If, upon further study, it is found that all degrees of intelligence occur at a constant level of sanity and that all degrees of sanity occur at a constant level of intelligence (8), then the two traits are for all practical purposes independent. If this be true, then intelligence has no direct effect on sanity, and sanity has no direct effect on intelligence. High intelligence does not prevent anyone from becoming or being insane. Insanity does not cause a person to lose his intelligence. Low intelligence may be associated with a high degree of sanity. To borrow a phrase from factor analysis, intelligence and sanity are placed orthogonally to each other in a random sampling of the population. Nevertheless, intelligence and sanity profoundly influence all adjustments simultaneously and differentially. Both may be positively correlated with tests but uncorrelated with each other. It is probable that scatter varies in extent and nature with the combination of the two traits in each individual.

Objective evidence for such a conclusion has so far not been provided. The results of scatter studies are contradictory and confusing. The limitations of tests in the detection of mental disturbances, discussed by Magaret and Wright (20), may not be inherent in scatter analysis as such but in the methods and theories employed in its measurement. Psychometric science may successfully catch up with the dictates of common sense if its scope of inquiry is extended without prejudice to all basic qualities of the personality.

The number of traits directly measured by tests are more than two. They are probably different from those used in our simplified example even though these two variables form the central point of interest of nearly all investigators of scatter. The value of mental tests will be greatly enhanced as soon as we stop deciding beforehand what we wish to measure and confine ourselves to the empirical investigation of what is actually being measured.

In the process of developing a qualitative and quantitative system of scatter analysis certain precautions should be taken to obviate extreme swings of the pendulum from a rigid and faulty statistical base to the loose and vague hypothesizing such as is typical of the projective techniques. Scatter analysis may fail to satisfy the requirements of an objective diagnostic aid if certain theoretical pitfalls and technical difficulties are overlooked in its study. The following pages will be devoted to the discussion of the most salient problems associated with scatter analysis.

SCALE STANDARDIZATION AND SCATTER

Thorndike once said that whatever exists, exists in a certain amount. The numerous attempts at quantifying scattering attest to the accuracy and wisdom of his observation. If scatter is orderly (16, 26, 29) and represents some consistent traits in the examined individual, then such traits should ultimately be measurable through indices of response variability. Despite the discouraging results of scatter analysis on the Stanford-Binet (**11**, **17**, **27**), the process is being repeated on an even larger scale with the Wechsler-Bellevue tests.

The Wechsler-Bellevue Scale has several important advantages over the Stanford-Binet Scale. It measures eleven separate abilities in fairly well-graded sub-scales. It has intra-test homogeneity and inter-test heterogeneity. This cannot be said of the Stanford-Binet Scale. In fact, the latter test measures with fair consistency only one type of ability, which is closely related to Thurstone's (**33**) verbal group factor.

According to McNemar (**21**), all items of the Stanford Scale are saturated with one general factor. This factor is sufficient to account for most of the inter-correlations between items except at four specified age levels (2, 2½, 6, 18). McNemar also finds a motor factor and a memory factor which may be responsible for some test variability at certain age levels. He warns, however, that the sampling of tests in each of these areas is highly unreliable, and therefore no sweeping pronouncements concerning special defects may be made on the basis of scatter. McNemar's analysis clearly indicates that the Stanford-Binet Scale lacks the much desired inter-test heterogeneity. It would have been a superior diagnostic tool if substantial samplings of many different abilities had been included at all levels.

The high correlations of most items with the general factor may be said to be spurious because of the biased ability sampling of the scale. What McNemar calls a general factor may in reality be a combination of several factors. The differences in common factor saturation between the various items are sufficiently great and consistent to posit the existence of a verbal group factor in addition to the general factor. This hidden group factor, though as constant and stable as the general factor, is responsible for a great deal of scattering on the Stanford-Binet Scale in abnormal cases.

From our experiences with the Stanford-Binet Scale it may be concluded that attempts to assemble test items which are highly correlated with each other and with the whole scale should be abandoned in favor of less correlated but clinically more valuable items. A saturation coefficient of over .60 with any factor should be viewed with considerable suspicion, as it probably represents a statistical distortion produced by improper test samplings or by criterion contamination.

If the theory of multiple traits (**16**, **33**) holds, then high correlations between tests and factors or traits are little more than standardization artifacts. For the relationships between tests and traits, whether general or specific, are mediocre at best.

Since the centroid verbal factor is so prominent and extensive in the Stanford-Binet Scale, it is likely to misrepresent the mental levels of all intelligent persons who are inferior in this dynamic trait. This is one of the most serious shortcomings of the Scale. The I.Q.'s obtained from such a scale should be used with utmost caution in all diagnostic work with problem children.

High saturation of tests with factors supports the frequently recurring idea that there are pure tests of this and that factor (8, 37). Eysenck, for example, uses the Raven Matrix test as a measure of intelligence because it is supposed to be a test of pure "*g*." This method of measuring intelligence contradicts Eysenck's very poignant argument that each test is the result of at least four factors, one general, one centroid, one specific, and one error factor. The theory of manifold causation is so reasonable and pertinent that it should relegate the idea of pure factorial tests to the realm of statistical myths. The reviewer suspects that the Raven Matrix test measures several important group factors in addition to the general factor. To find that the Matrix test is a relatively pure test of "*g*" must surely be due to faulty experimental design. The same may be said of tests which are supposed to be pure measures of some group factor (37). The existence of pure tests of either general or group factors cannot be reconciled with psychological realities in clinical practice.

One of the greatest, and as yet unsolved, problems of the experimental laboratory and clinic alike is the coordinate interpretation of the results from many different tests. Rapaport (26) and Eysenck (8) treat the large number of different tests and techniques without making direct comparisons between them. This procedural atomism not only breaks up the diagnostic study into artificial and overlapping sections but also retards the final determination of what the various tests measure and how well they measure what they do. When the Bellevue and Babcock tests (26) differentiate between neurotics and schizophrenics, they may not necessarily do it in the same manner. The factors involved may not be identical or even similar. When tests of vocabulary, social history, persistence, choline esterase, and body build differentiate between Eysenck's (8) hysterics and dysthymics, it does not follow that all these methods are good and valid tests of the same temperamental factor as Eysenck seems to imply. The way to find out what is being measured by all these tests is to administer them to the same experimental population and to a control group and then subject them to an over-all factor analysis. In this manner, the similarity or dissimilarity of the analytic criteria would soon become apparent. It would also facilitate the discovery of several differential signs based not on the nature of the testing procedure but on the nature of the psychic traits or genotypes measured.

FOURFOLD ANALYSIS OF TESTS

Psychometric batteries composed of a large variety of independently scaled, homogeneous abilities yielding separate scores are conducive to a fourfold analysis of test variability.

First, they permit the comparison of one individual in any ability with the norm of many other persons of the same age in that ability. This inter-personal index is similar to the I.Q. except that reliable scores obtained from several discrete test functions increase the analytic possibilities. They provide a multi-dimensional system of comparison.

Second, homogeneous subscales measure the intra-personal organization of functions which are directly and intimately related to dynamic personality traits. The discrepancies between scores within each individual record are of great diagnostic importance (**4**, **5**, **8**, **16**, **17**, **23**, **26**, **28**, **29**, **30**, **31**, **34**).

Third, the response variations of a person within each subscale may be used in the interpretation of test results. Two persons obtaining identical results on an information test, for example, may have arrived at this score in a totally different manner. Indeed, identical score patterns within a scale are a rare exception in records with identical final achievements. Eysenck (**8**) attempted to measure intra-test scatter in relation to the Matrix test. Because of the method he used and because of the lack of direct inter-test comparisons his results were negative as far as differentiation between his diagnostic groups was concerned. The reviewer (**16**) believes that intra-test scatter may contribute to the understanding of personality dynamics. It is to some extent related to inter-test discrepancies. It defines the nature of inter-test variability.

Fourth, a test response may be analyzed in regard to several aspects which have so far not received sufficient objective evaluation. They are, however, given some attention in Rapaport's (**26**) and Schafer's (**29**) writings. Reaction times, rate of production, amount of output, contents of response, form of response, nature and number of errors preceding the solution, and projective qualities deriving from emotionally tinged, personal experiences are important factors in all tests. These facets of a response undoubtedly influence the clinician's judgment even though they receive no methodical statistical analysis. That such an analysis should ultimately be made to keep diagnostic inferences within reasonable bounds is self-evident. It is precisely these unobjectified and intangible qualities of a response that lend themselves to the making of far-fetched and exaggerated claims. Clinical rationalizations may be kept to a minimum through the adoption of multiple scoring techniques which, if necessary, overlap some or all of the subscales of a battery. The appraisal of characteristic and psychologically uniform response deviations appearing intermittently throughout the battery may lead to the isolation of a number of helpful diagnostic signs.

Like life itself, tests are neither fully structured nor unstructured. Projective elements are liable to occur on all tests. A vocabulary test, properly standardized, may be equally as projective as the Rorschach test (18). Furthermore, the measurement of projective qualities by means of vocabulary and other psychometric tests is likely to be more precise and more meaningful than are similar measurements from so-called unstructured tests. A change in the theoretical foundation of the Binet type of test will produce more satisfactory diagnostic and therapeutic indicators than does its replacement by tests which elude scientific treatment.

The Wechsler-Bellevue Scale (34) or any similarly constructed battery (14, 15) may be a valuable tool of multiple response analysis if the necessary precautions are applied to the theoretical and technical problems of scatter measurement. The potentialities inherent in such an analysis are great and numerous. Variability studies are certainly as important as the final test scores and more informative than intelligence quotients.

A good psychometric scale should, indeed, be standardized for the various types of scatter in addition to providing achievement norms. Before this is done, psychologists will have to be satisfied with cluster analysis as an after-thought rather than as a primary diagnostic aid.

EXTERNAL CRITERIA OF VALIDATION

Scatter is more often observed in mentally ill or disordered persons than in normals. Since most scatter studies come from psychiatric clinics and mental hospitals, it is natural to find them evaluated in terms of psychiatric disease entities. Two objections may be raised to such a procedure. First, attendance at a clinic or commitment to a hospital does not make a person abnormal and, vice versa, not being a psychiatric patient does not make a person normal. Second, psychiatric classifications of mental illness are mono-dimensional and therefore do not fully agree with all the facts in the given case. Psychiatrists generally consider their present-day nosology outdated and unsatisfactory. A great deal of research will have to be done to objectify the syndromes associated with mental illness and personality disorganization. The very existence of disease unities such as are postulated in psychiatric diagnoses awaits experimental verification.

The criteria differentiating one illness from another are vague and fluid. The diagnosis therefore depends largely on the school and subjective bias of the individual psychiatrist. Some students of scatter analysis (**8**, **16**, **26**, **29**, **34**) are fully aware of the implications inherent in the use of comparative standards of low or dubious validity. There may be substantial agreement between psychometric findings and psychiatric evaluations outside the reference frame of official nomenclatures. However, most scatter analyses are based directly on official diagnostic categories. The outcome of such studies hinges as much on the objectivity of the psychiatric grouping as it does on psychometric analysis.

This dilemma may be overcome either by abandoning the practice of validating scatter against psychiatric diagnoses as now conceived and employed, or by obtaining from the psychiatrist a comprehensive evaluation of the adjustment pattern of each patient. Another method of improving validation criteria is to give greater weight to the intrinsic relationships of test scores obtained from the experiment. Chemists, physicists, and astronomers depend almost exclusively on criteria of internal consistency set up as a result of well-planned laboratory procedures. Objective psychology has reached a stage at which criteria of internal consistency may be of considerable service. For example, the discrepancy between drawing and arithmetic scores is a more valid criterion of mental adjustment than any extraneous criterion except the actual life developments of the individual case. These developments should be checked and studied by the experimenter directly without complicating the issue by seeking opinions and impressions of inferior value. There may, in fact, be a closer agreement between psychometric test criteria and living realities than is generally assumed to be the case. Psychologists will make their greatest contribution to psychiatry when they emancipate themselves from current psychiatric systems of diagnoses and evolve a rigorous and multiple system of comparison between integrated and unintegrated personalities.

REFERENCE POINTS

The phenomenon of response scatter is a universal one. It has many degrees and many qualitative variations. It occurs in persons of all ages and types. The general assumption is that scatter is greater in those with abnormal than in those with normal personalities. This assumption will be found to be correct when tests are standardized for scatter on the basis of criteria of internal consistency. One of the simplest and most direct measures of scatter is the range ratio. It is obtained by dividing the lowest score from a large number of subscales by the highest score from the same tests (**16**). The range ratio of the

Wechsler-Bellevue test is considerably smaller in all abnormal groups than it is in the normal population. It indicates that disturbed and unorganized people are more variable, less dependable, and less compact in most of their adjustments (**16**). The range ratio is an objective measure of the degree of mental fluctuations and general variability. It does not distinguish between different types of scatter within each record.

When attempts are made to determine the scatter clusters between or within tests, the choice of a reference point from which scatter is measured becomes important. The psychological and statistical soundness of this reference point governs the soundness of the final scatter results. It must have statistical stability and psychological relevance. In other words, it must be clinically valid and statistically reliable. Its choice is frequently a function of the personality theory to which the experimenter subscribes.

Three types of reference points have so far been used in the study of test scattering:

1. Deviations of test scores from the stable score of a test which is usually believed to be a good test of original intelligence. Babcock's vocabulary scatter (**1**, **2**) and Eysenck's Matrix test scatter (**8**) are two examples of this type.

2. Deviations of scores from the mean of all tests or of a group of tests. This is one of the most popular methods of scatter analysis. It has been used by Wechsler (**34**), Rapaport (**26**), Schafer (**29**), and many others.

3. Factor or cluster analysis in which the mean of one cluster is compared with the means of other clusters of tests. Thustone's (**33**) group factors and Jastak's (**16**) altitude scatter belong in this category.

Let us scrutinize the practical value of these three procedures in the light of clinical experience and experimental fact.

Vocabulary Scatter

The advantages of comparing vocabulary scores with other test scores are two. The vocabulary yields one of the most stable rating used in psychometric examinations. It is also relatively invulnerable to the presence of disorders and disturbances such as are observed in psychotic persons. This refractoriness of the vocabulary score has been noted very soon after the introduction of tests as a method of measuring intelligence. Wells (**35**) was the first one to use it in a systematic manner for purposes of measuring the variability of mental patients. Babcock (**3**, **4**) has elaborated the relationship between vocabulary and other tests into her well-known "level-efficiency" theory. The vocabulary test, being uninfluenced by mental disorganization, serves as an index of native capacity, while the highly sensitive efficiency tests, being reduced by mental disturbances, serve as measures of sanity. The relationships between the two types of tests may thus be employed to determine the degree of deterioration. Babcock (**1**, **2**) has provided experimental evidence for the partial correctness of her theory by testing it on patients suffering from general paresis and dementia praecox.

More recently, Rapaport (**26**) confirmed the diagnostic value of vocabulary scatter by using Babcock's tests in clinical examinations. He states "These statistical data then substantiate the contention that Vocabulary is the least variable and one of the best retained of the subtests; thus it might well serve as an indicator of the original intelligence level, and as a standard of comparison for the other subtests." The two premises underlying this method of scatter analysis are that vocabulary is a good test of intelligence and that its score is insignificantly depressed by mental illness. Though vocabulary has been shown to decline somewhat with age (**34**) and to decrease significantly in cases of organic deterioration (**6**, **7**), its relative invulnerability has been sufficiently documented by actual test throughout the history of clinical psychology. The stability postulate of vocabulary scatter might, with some reservations, be clinically acceptable.

The assumption that the vocabulary is a superior test of intelligence depends on the definition of intelligence and on actual correlations between intelligence and vocabulary. Since a general theory of intelligence has so far not been evolved, the postulate that vocabulary gives valid measures of intelligence remains experimentally unverified. Jastak (**16**) finds that the vocabulary is no better a test of native capacity than is any other well-standardized test. If this is true, then vocabulary scatter should prove quite unsatisfactory and even misleading in a large number of mental patients. It is imperative that the reliability of a score not be confused with its validity. The Wide Range Reading Test (**13**) is even more stable and less vulnerable in mental examinations than is the vocabulary. Yet we would not dare suggest that oral reading is a valid test of pre-psychotic intelligence because no test can be credited with such unusual powers, no matter how stable its score may be. This criticism applies in even greater degree to Raven's Matrix test used by Eysenck (**8**) as a measure of intelligence.

If a person has a defective vocabulary despite good intelligence, then the difference between vocabulary and efficiency tests may be positive or may indicate a favorable degree of personality organization even though the person is severely disorganized. Furthermore, an individual may have an efficiency score of 120 and a vocabulary score of 75 before he becomes psychotic. During psychosis the efficiency score drops 40 points (from 120 to 80), whereas his vocabulary score remains

75. In such cases (not at all rare in clinical practice) the vocabulary scatter is useless as a measure of efficiency for the simple reason that it is not an accurate measure of original intelligence. It is incapable of revealing the decline in efficiency which has actually occurred.

Vocabulary scatter is valuable only in a small number of cases such as are seen in private clinics and other institutions catering to a highly select group of patients. It fails to distinguish between integrated and unintegrated behavior in the majority of cases treated in clinics and hospitals which draw their patients from all segments of the community. The reviewer (**14**) has found that the magnitude and the distribution of vocabulary scores among mental patients are similar to those of normal persons. This would indicate that vocabulary is indeed uninfluenced by mental disturbances. This does not, however, entitle us to conclude that the vocabulary test always furnishes an accurate classification of original intelligence. Mental levels based on vocabulary scores are accurate and valid in not more than 10 per cent of mental patients. Clinicians will, therefore, do well to use the vocabulary test with utmost caution in the determination of intellectual level and also in the determination of efficiency by means of vocabulary scatter.

Babcock (**3, 4**) has advanced psychometric theory by adding the efficiency dimension to the intelligence quotient. This extension of the usefulness of tests is valuable but it does not go far enough. The human personality is multi-dimensional. The most complete system of psychometric interpretation must therefore be manifold. As Rapaport (**26**) puts it "... Vocabulary Scatter ... is of little value in gauging the general spread of scatter over the scattergram. Our approach to scatter analysis is to consider the relation of all subtests scores to each other. Considering scores in relation to the vocabulary level is only one aspect of this approach." A multi-dimensional approach is not limited to one rigid reference point, but employs many reference points on the basis of their empirical value in test analysis. The recognition of this important tenet will obviate the elevation of one test to the pedestal of an exclusive reference point (**2, 8**) since it limits the possibilities of meaningful scatter analysis through the study of many other relationships.

Mean Scatter

In appraising the diagnostic and clinical features of the Bellevue Scale, Wechsler (**34**) has found that the deviations of the eleven subtest scores from the means of all scores bear certain consistent relationships to the disorders of psychiatric patients. Wechsler uses differences of individual test scores from the means. Rapaport (**26**) and Schafer

(**28, 29, 30**) employ a similar method of measuring scatter. This technique is, of course, more subtle and more versatile than vocabulary scatter. It comes close to a multi-dimensional system and satisfies to some extent the criterion of psychological polivalence of test scores. It should prove serviceable if the mean of all scores meets the validity requirements of analysis. The mean of all scores is undoubtedly a very stable reference point. Whether it is psychologically relevant is far less certain. Its relative constancy is an asset, its psychological ambiguity is a liability. The mean of all scores often acts like an unwanted middleman. It cuts the real discrepancies and disregards the direct relationships between tests. The difference of 9 points between the weighted score of 12 on the vocabulary test and the weighted score of 3 on the arithmetic test may be more meaningful than are the two scores of $+4$ and -5 from a mean of 8 points.

The difficulties of mean scattering increase when the mean is close to the extremes of the distribution of individual weights. In such cases, it is a highly contaminated statistic. It neutralizes the very deviations which are of diagnostic importance. An acutely psychotic or seriously disorganized individual may have a low mean score as a result of his personality defects. The actual discrepancies from the mean, whether cumulated or not, may give no evidence of a disturbance when in reality it is very severe. The more generalized the mental disorder the less efficiently does the mean differentiate between normal and abnormal functioning. The general thesis may be adopted that the use of any reference point for the measurement of a trait is impractical if the trait in question has a variable effect on that reference point. If the mean of all tests, like the vocabulary, were unaffected by personality impairment, it could well serve as a reference axis of scattergrams. Since this requirement is never satisfied in clinical studies, the mean of many tests may not be regarded as a psychologically adequate pivot about which scatter varies in a consistent manner. Like the vocabulary test, it has a great deal of reliability but little validity for the accurate appraisal of scatter. The vocabulary score is relatively uninfluenced by mental disorganization, but it is not a valid measure of original intelligence. The mean score is an invalid index of original intelligence because it is influenced by mental illness.

A third flaw of the mean is that the various types of mental conditions have different effects on it in different persons. It may thus give an accurate picture of a trait in one individual, but obscure the very same trait in another individual.

Cluster Analysis

It is infrequently realized that clinicians and factor analysts are doing the same thing in different ways. When Babcock (3) decided to divide her efficiency battery into motor, learning, and memory tests, she was engaging in intuitive factor analysis. When Wechsler (34) grouped the subtests of the Bellevue Scale into verbal and performance parts, he assumed that the two types of tests measured different functions. When Rabin (25) divided the scores of one cluster of tests by the scores of another cluster to differentiate between schizophrenia and manic-depressive psychosis, he applied a clinical method of factor analysis. When Rapaport (26) eliminated the digit span and arithmetic tests from his verbal scatter measures, he employed a legitimate method of cluster analysis which could have been readily confirmed by Thurstone's centroid method of factor analysis.

Factor analysis (8, 16, 33, 37) accomplishes by statistical means in a relatively short time what the clinician discovers after many years of hard-won experience. Even then, the clinicians' hunches may be incomplete and unrefined in comparison with the differentiations supplied by factor analysis. That is why Thurstone's group factors may be considered another form of scatter analysis.

Factor analysis has the advantage over other systems in that it reduces the almost infinite complexity of test interrelationships to a smaller number of psychologically pertinent unities. It directs our attention to the weighty fact that scatter analysis has as many facets as there are factors. Scatter may be large in some tests and small in others depending on the personality structure of the examined individual. Unless we learn to differentiate between types of scatter within the same test battery, we are likely to end up with few or no positive signs despite Herculean efforts. When scatter scores are averaged without regard to the clusters in which they occur, their diagnostic significance may disappear in the process of analysis.

The strong points of factor analysis may be summarized as follows:

1. It is an objective and efficient means of grouping psychologically homogeneous tests.
2. It provides a highly differentiated system of scatter analysis within the isolated clusters.
3. It supplies several stable and important reference points that are clinically meaningful.
4. It shows that one sub-test may belong to two or more clusters and that such tests should be used in the appraisal of more than one trait. This latter possibility is especially difficult to gauge by mere clinical impressions.

The clinical application of factor analysis has been limited in the past because of several shortcomings in the interpretation of factors.

1. The naming of factors after test contents does not satisfy the clinical psychologist who thinks more in terms of personality dynamics than in terms of test materials.

2. Factor analysis has not been able to provide factorial scores which are actually uncorrelated with each other. For example, the number factor and the verbal factor always yield scores that are positively correlated. The independence of these factors has not been empirically demonstrated up to this point.

3. Those who believe in group factors ignore the possible existence of one or more general factors which are responsible for the positive correlations found between group factor scores. Unless these general factors are accounted for and held constant, scatter analysis by factoring will be of little avail to the clinician. In fact, the general factor, properly measured, might be used as a stable and valid reference point against which the group factors are evaluated. Such a method was recently suggested by the reviewer (**16**).

4. The scope of factorial interpretations must be enlarged to include the entire repertory of personality traits. Groups factor may in reality represent personality or character traits which are independent of intellectual capacity. Factor analysis is more adequate in its statistical aspects than in matters of mental analysis. To improve its practical usefulness, the following methodological improvements are proposed: (a) the isolation and measurement of the general factor of intelligence; (b) the isolation and identification of group factors with non-intellectual character dynamics; (c) the study of the relationships of the general factor to group factors; (d) The empirical demonstration that the deviations of group factor scores from intelligence are independent of the degree of intelligence; and (e) the empirical demonstration that the deviations of the various group factor scores are independent of each other.

WEIGHTED SCORES AS MEASURES OF SCATTER

Scatter studies of the Wechsler-Bellevue Scale (**34**) are numerous and more objective than are those of any other test scale. However, most of the published studies suffer from a number of technical deficiencies which should here be pointed out. One of these is the almost universal use of weighted scores.

As is well-known, the raw score of each one of the 11 Bellevue subtests is converted into a weighted score to equate the achievements of different persons on the same test. The weighted score is in turn transmuted into an intelligence quotient. The weighted score and the intelligence quotient are statistically equivalent for they are both standard scores. Psychologically, they differ significantly because the weighted score disregards age whereas the quotient does not. The Wechsler quotient is thus a convenient method of keeping age relatively constant. Scatter patterns are more accurate and more meaningful when expressed in terms of quotients rather than in terms of weighted scores.

Rapaport's (26) 67 schizophrenics have a mean age of 31 years; his 33 depressed patients have a mean age of 49 years. The depressives are on the average 18 years older than are the schizophrenics. Rapaport finds that "No schizophrenic group, not even the Deteriorated Schizophrenics, show such a consistently great impairment of all Performance subtests as the two Depressive Psychosis groups. This scatter pattern is of crucial diagnostic significance for differentiating Depressive Psychoses from Schizophrenia."

If age is kept constant by the use of test quotients, his depressive groups fail to show the consistently great impairment of performance subtests. When all the precautions discussed in this paper are applied to Rapaport's cases, the average performance scatter score (deviation ratio from capacity) for his 44 most chronic and deteriorated schizophrenics is 82.9. The average deviation ratio for his 33 depressives is 82.2. The difference is insignificant. Both scores are significantly below the average (normal) scatter score of 87.4. Rapaport's deteriorated unclassified schizophrenics (N=7) have a scatter score of 74.7 which is below that (78) of his most disorganized depressives (N=8). Rapaport's crucial scatter differences are more a function of the age differences between the diagnostic groups than they are of the groups themselves. Foster (9) has shown that when weighted scores are corrected for age, the originally diagnostic scattergram turns out to be an artifact. The test discrepancies either become smaller or disappear when age is kept constant.

It may be added that the performance deviation ratios of Rapaport's groups are generally high and the discrepancies rather small. They approach the norm of the population at large much more closely than do the ratios of hospital patients studied by the reviewer. This may be explained in two ways. Either Rapaport's cases were less disorganized than are psychiatric patients generally. Or, the diminished scatter is due to the fact that his patients were of high average and superior intelligence. High capacity as such is not a deterrent in the study of scatter. However, our experience indicates that the Bellevue Scale is least satisfactory at the extremes of its ability range. It does not give the superior person sufficient opportunity to demonstrate his true superiority. Nor does it give the stupid person a full chance to show his real level of stupidity. This telescoping effect of the Bellevue Scale is a distinct weakness in all investigations of scatter. Patients are prevented from scattering because of the limitations of the scale at both ends.

Any scatter analysis of the tests of the Bellevue Scale by means of weighted scores is of dubious value, unless the experimental or clinical groups are homogeneous in age or unless test quotients are used to allow for age differences between members of such groups.

SEX DIFFERENCES

It was seen that age differences may produce typical scattergrams. These patterns are of little value in diagnosis if not corrected for age. Another factor which is likely to disturb the clinical significance of subtest analysis is sex. Eysenck's study of test variability (8) is one of the most satisfactory of all those published in that he always separates the two sexes.

As long as psychometric ratings are expressed in averages from many different abilities, the differences between males and females are neutralized. This is especially true if the battery consists of an equal number of tests favoring males and females. As soon as the averages are broken up into subcomponents, significant sex differences may appear in some of the clusters. Scatter analysis will then show disorganization where it does not exist. Or, it may show normal organization in disturbed individuals. Sex differences may not be ignored in the interpretation of various types of scatter. On the Bellevue test, males tend to be better in information, picture completion, object assembly, arithmetic, and digit span. Females tend to be better in vocabulary, symbol substitution, comprehension, picture arrangement, and block designs.

Vocabulary scatter may, in the light of these differences, mean something different in case of women than in the case of men. If a predominantly female cluster of tests happens to be used in the computation of scatter scores, it may give a distorted picture of personality functioning just because the patient is a man or a woman.

Future psychometric scales, if standardized for scatter, will have to be equated for sex differences.

ARITHMETICAL DIFFERENCES AS MEASURES OF SCATTER

It is common practice to express test discrepancies in terms of differences between scores. Thus vocabulary scatter (**1**, **2**, **26**) is usually measured by subtracting the vocabulary score from the efficiency score. If the difference is positive or small, the subject is assumed to be normal. If the difference is a large negative one, the patient is thought to be deteriorated. Weighted and sigma scores are used in the calculation of differences, presumably to render them comparable. In the mean scatter method, the subtest score is subtracted from the mean. Babcock (**1**, **2**), Wechsler (34), and Rapaport (26) use the method of differences between weighted or standard scores. Rapaport (26) mentions the necessity of dealing with positive and negative numbers as bothersome, but this is not the real difficulty of the method.

Wechsler (34) points out correctly that identical differences have different meanings at different achievement levels. Thus a difference of 3 points at the weighted score level of 3 to 6 is twice as great as the difference of 3 points at the 9 to 12 level. If this important fact is ignored, the findings of scatter analysis from the method of differences are incomparable. It must be stressed that differences in terms of sigma scores do nothing to alleviate this problem.

Differences are positively correlated with the magnitude of the subtrahend. Wechsler (34) overcomes this difficulty by dividing the total weighted scores of ten tests by 40. The quotient thus obtained indicates the significance of the differences between scores at different test levels. When the sum of the weighted scores is 120, a difference of 3 points (120/40) is significant. When the sum of the 10 weighted scores is 60, a difference of only $1\frac{1}{2}$ points (60/40) is significant.

Jastak (14, 15, 16) has achieved the same results by dividing directly the smaller standard score by the larger. One of the characteristics of ratios is that they are uncorrelated with the denominator used in their calculation. The best known example illustrating this principle is the intelligence quotient. The I.Q. is by definition uncorrelated with the chronological age. If an index of brightness had been proposed in terms of the plus-minus differences between mental and chronological age, it would have the same unwieldy and misleading properties typical of most scatter measures now in use.

The sums of differences between scores for two persons are not comparable unless both persons have identical subtrahends. The mean difference scatter of deteriorated schizophrenics is not comparable with that of depressives unless both groups have identical means.

Ratios are generally more useful in the study of the relationships between tests than are differences (10). They may, for example, be conveniently employed in measuring the correlational residuals of group factors when the general factor is kept constant (14, 15, 16). Dividing the group factor scores by the general factor scores will accomplish this end. The deviation ratios from capacity of each subtest of a homogeneous cluster are positively correlated with each other. On the other hand, the size of the deviation ratios of a cluster from capacity is uncorrelated with capacity. Furthermore, the deviation ratios of one cluster may be shown to be uncorrelated with the deviation ratios of another cluster. When the personality trait representing a group factor stands in a compensatory relation to intelligence, as happens in some highly selected vocational groups, the correlation between deviation ratios of one cluster and intelligence (or another cluster) may actually be negative. In random samplings, the correlations between general and group factors approach zero.

That human adjustments are governed by compensatory relationships between dynamic traits is more than a popular belief. Scatter and cluster analysis may become appropriate vehicles for demonstrating that beyond the correlational realm of measured abilities there exist traits or factors which are either uncorrelated or which, under certain conditions, may take up opposite poles in relation to over-all personality functioning.

The scientific proof of trait compensation devolves upon one important contingency. Select groups may not be treated as self-sufficient units, but should be compared with the norms of a broader universe—the total population. The experimental controls must be random in regard to all personality traits which characterize people because all these traits share in the production of scatter on tests. The original sampling must be random at all age levels and for both sexes.

PSYCHOSIS, COOPERATION, AND SCATTER

Eysenck (**8**), Rapaport (**26**), Jastak and Vik (**15**) speculate about the possible relationship between psychosis and psychometric evidence of disorganization. The consensus of opinion seems to be that psychometric disorganization is more basic and more far-reaching than is the fact of acute psychosis. Abnormal scattergrams may precede the onset of psychosis by many years and may continue unchanged following remission. Many psychiatric remissions are surface phenomena unaccompanied by deep-seated changes in personality organization. The personality deviations measured by tests may serve as the background for the chronic or intermittent exacerbations in the form of delusions, hallucinations, and acute psychotic states.

The correlation between degrees and types of scatter and psychosis is not high, as Wittman (**38**) and Brody (**6**) point out. There is no valid reason for assuming that such correlation should be greater than moderate at the very best. Attempts to prove that psychometric disorganization and psychosis are highly correlated is a futile undertaking. Whenever high correlations between psychosis and psychometric scatter are reported, they may be considered the result of ill-designed experiments. A psychosis is not infrequently an acute and temporary condition superimposed upon a relatively normal personality. To measure scatter and disorganization in psychotic and deviant persons, the scatter of normal people must be known. The judicious use of internal criteria of consistency will demonstrate that those usually considered normal may not necessarily be so, and obversely, those considered abnormal may not be so either.

Wittman (36) expresses the view that psychometric scatter as measured by Babcock's method may be a function of test cooperation on the part of the subject. She is seconded in this view by Brody (6). Acceptance of this opinion depends on the definition of cooperation. Inability to cooperate is usually a part symptom of the mental disturbance and therefore need not cause serious concern as a major factor in scattering. Great skill and patience are required to administer tests to disturbed individuals. The results of such tests are fully as valid as those obtained from entirely cooperative persons. When the patient is unwilling to cooperate, the test is incomplete and scatter analysis impossible. There is little else to report except the fact of refusal and the circumstances under which it occurs. The reviewer is inclined to the view that the degree of cooperation is a very minor factor in the production of test scatter.

Psychometric results are invalid as measures of basic personality traits only in cases of profound stupor, seriously reduced consciousness, and complete mental confusion.

SUMMARY

Scatter analysis is a central problem in all psychometric examinations. Objective comparisons may be inter-personal and intra-personal. Inter-individual norms delimit the standing of a person in any ability or trait in comparison with other persons in that ability or trait. Intra-individual norms may be established by the study of inter-test discrepancies and intra-test response patterns. Inter-test discrepancies require a systematic survey of the relationships between many tests and abilities within the same individual. The method of choice is factor analysis complemented by clinical insight and a personalistic approach.

The reference point of scatter analysis should be provided by a psychologically homogeneous and statistically stable trait. A general factor, accurately measured and operationally defined, is the most suitable for this purpose. Neither vocabulary scatter nor mean scatter, though they constitute invulnerable and stable statistics, is clinically applicable to the study of personality traits. The vocabulary test may not be considered an accurate and valid measure of pre-psychotic intelligence.

The mean of all scores, on the other hand, is often a highly contaminated value which reduces the sought-after variations in a variable manner.

Psychometric scales should be standardized for scatter as well as for achievement scores in different abilities. Standardization samplings should be random in regard to all measurable personality traits. External criteria of validation should either be improved or supplanted by criteria of internal consistency determined in the course of the experiment.

The use of weighted scores or similar measures uncorrected for age and sex interferes with the objective appraisal of clinically valuable scatter records. The method of differences between scores and a reference point should be replaced by the method of ratios of one score to the other. The method of ratios eliminates positive correlations between the magnitude of the reference point and the size of the discrepancies from it. It makes the deviations comparable at all levels of achievement.

The most far-reaching change in regard to scatter analysis is the extension of the value of psychometric tests to the field of general personality measurement. The restrictive pars-pro-toto theory of intelligence should be abandoned. Furthermore, the idea that tests of pure factors can be assembled is contrary to the principles of personality unity and simultaneity. It may be safely assumed that any ability whatever is the function of at least four types of factors: general, centroid, specific, and accidental.

BIBLIOGRAPHY

1. Babcock, Harriet. An experiment in the measurement of mental deterioration. *Arch. Psychol. N. Y.*, 1930, No. 117.
2. Babcock, Harriet. *Dementia praecox: a psychological study.* Lancaster, Pa.: Science Press, 1933.
3. Babcock, Harriet. Personality and efficiency of mental functioning. *Amer. J. Orthopsychiat.*, 1940, **10**, 527–531.
4. Babcock, Harriet. The level-efficiency theory of intelligence. *J. Psychol.*, 1941, **11**, 261–270.
5. Bijou, S. W. The psychometric pattern approach as an aid to clinical analysis—a review. *Amer. J. ment. Def.*, 1942, **46**, 354–362.
6. Brody, M. B. A survey of the results of intelligence tests in psychosis. *Brit. J. med. Psychol.*, 1942, **19**, 225–261.
7. Davidson, M. Studies in the application of mental tests to psychotic patients. *Brit. J. med. Psychol.*, 1939, **18**, 44–52.
8. Eysenck, H. J. *Dimensions of personality.* London: Kegan Paul, 1948.
9. Foster, A. Age and the Wechsler-Bellevue scattergraph. *J. clin. Psychol.*, 1947, **3**, 396–397.
10. Hamlin, R., & Abel, T. M. Test patterns of mental defectives skilled in-

weaving. *J. appl. Psychol.*, 1938, **22**, 385–389.

11. Harris, A. J., & Shakow, D. The clinical significance of numerical measures of scatter on the Stanford-Binet. *Psychol. Bull.*, 1937, **34**, 134–150.
12. Jastak, J. *Variability of psychometric performances in mental diagnosis.* New York: Privately published, 1934.
13. Jastak, J. School test patterns of clinic children. *Del. State med. J.*, 1938, **10**, 108–111.
14. Jastak, J., & Gordon, G. J. Verbal polarity in dementia precox. *Del. State med. J.*, 1946, **18**, 131–136.
15. Jastak, J., & Vik, E. S. The objective measurement of reality perceptions in dementia precox. *Del. State med. J.*, 1947, **19**, 98–102.
16. Jastak, J. A plan for the objective measurement of character. *J. clin. Psychol.*, 1948, **4**, 170–178.
17. Kendig, Isabelle, & Richmond, Winifred V. *Psychological studies in dementia praecox.* Ann Arbor, Mich.: Edwards Bros., 1940.
18. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* New York: World Book Co., 1946.
19. Lorr, M., & Meister, R. K. The concept of scatter in the light of mental test theory. *Educ. psychol. Meas.*, 1941, **1**, 303–309.
20. Magaret, Ann, & Wright, C. Limitations in the use of intelligence test performance to detect mental disturbance. *J. appl. Psychol.*, 1943, **27**, 387–398.
21. McNemar, Q. *The revision of the Stanford-Binet Scale.* Boston: Houghton Mifflin, 1942.
22. Piotrowski, Z. A. Objective signs of invalidity of Stanford-Binet tests. *Psychiat. Quart.*, 1937, **11**, 623–636.
23. Piotrowski, Z. A. The test behavior of schizophrenic children. *Proc. Amer. Ass. Stud. ment. Def.*, 1937, **42**, 78–90.
24. Rabin, A. I. Test score patterns in schizophrenia and non-psychotic states. *J. Psychol.*, 1941, **12**, 91–100.
25. Rabin, A. I. Differentiating psychometric patterns in schizophrenia and manic-depressive psychosis. *J. abnorm. soc. Psychol.*, 1942, **37**, 270–272.
26. Rapaport, D., *et al. Diagnostic psychological testing*, Vol. I. Chicago: Year Book Publishers, 1945.
27. Roe, Anne, & Shakow, D. Intelligence in mental disorder. *Ann. N. Y. Acad. Sci.*, 1942, **42**, 361–490.
28. Schafer, R. The significance of scatter in research and practice of clinical psychology. *J. Psychol.*, 1944, **18**, 119–124.
29. Schafer, R. *The clinical application of psychological tests.* New York: International Univ. Press, 1948.
30. Schafer, R., & Rapaport, D. The scatter in diagnostic intelligence testing. *Character & Pers.*, 1944, **12**, 275–284.
31. Shipley, W. C., & Burlingame, C. C. A convenient self-administering scale for measuring intellectual impairment in psychotics. *Amer. J. Psychiat.*, 1941, **97**, 1313–1324.
32. Terman, L. M., & Merrill, M. A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
33. Thurstone, L. L. *Multiple factor analysis.* Chicago: Univ. Chicago Press, 1947.
34. Wechsler, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.
35. Wells, F. L. *Mental tests in clinical practice.* Yonkers, N. Y.: World Book Co., 1927.
36. Wittman, P. The Babcock deterioration test in state hospital practice. *J. abnorm. soc. Psychol.*, 1933, **28**, 70–83.
37. Wolfle, D. Factor analysis to 1940. *Psychometr. Monogr.*, 1940, **3**, 1–69.

Received January 12, 1949.

NUMERICAL TRANSFORMATIONS IN THE ANALYSIS OF EXPERIMENTAL DATA[1]

C. G. MUELLER[2]
Psychological Laboratory, Columbia University

This paper is concerned with some problems that arise in the analysis of experimental data and, in particular, with a set of operations frequently useful in such analysis, the numerical transformation. The term transformation, as here used, refers to the operations by which one set of numbers or scores is changed into another set. The procedures of computing logarithms or reciprocals of scores are familiar examples. In what follows we shall consider some of the reasons for converting a set of data into another form and discuss some of the problems that may arise from the application of such a conversion.

Transformations are used most frequently in three situations: (a) the situation in which a quantitative theoretical statement or prediction is reduced to a simpler form so as to permit a more convenient inspection of the agreement between data and theory; (b) the situation in which we seek an empirical equation to describe a set of data; and (c) the situation in which one frequency function is changed to another to allow the application of more efficient statistical tests.

TRANSFORMATIONS TO TEST THEORY

The reduction of a quantitative theoretical statement to a simple convenient form for test often begins with an attempt to express the relationship in linear form. [It is important to note that many expressions do not permit such manipulation (see Lipka, **31**).]

1. Consider a theory which demands that, to produce a threshold effect, the product of the intensity, I, of a light stimulus and its duration, t, be a linear function of the duration (**25**); that is

$$It = a + bt, \tag{1}$$

where a and b are constants. In such a case, a plot of the product, It, against t should give a straight line with a slope of b and an intercept constant a. Since equation [1] may also be expressed in the form

$$I = b + \left(\frac{a}{t}\right), \tag{2}$$

equation [2] presents an alternative way of testing the agreement between ex-

---

[1] Prepared under Project No. NR142-404; Contract N6onr-271, Task Order IX, between Columbia University and the Office of Naval Research, U. S. Navy.

[2] The writer wishes to express his appreciation to Prof. C. H. Graham for many critical suggestions as to the form and content of this paper. He is also grateful to Professors H. E. Garrett and Joseph Zubin for their comments on the manuscript.

perimental data and theory: a plot of I against the reciprocal of t should give a straight line with a slope of a and an intercept constant of b. Under these circumstances a reciprocal transformation of t might be used in analysis.

2. As another example, suppose there exists an hypothesis of visual figural after-effects of the following sort.[3] Assume that any symmetrical inspection figure viewed in a given position builds up a region of depolarization, and the extent of the area of depolarization for a given figure varies as a function of the exposure of the inspection figure. With cessation of stimulation by the inspection figure, the area of depolarization decreases in size as a function of time after inspection. The presentation of a test stimulus after removal of the inspection figure involves the interaction of the depolarization due to the test figure with that remaining from the activity of the inspection figure in such a way that the resultant depolarization at a point is the sum of the two depolarizations at that point. The position of an inflection point, L, in the resultant distribution of depolarization is a function of the distribution set up by the inspection figure, and the point of inflection, L, is assumed to vary with time after inspection according to the following equation

$$\frac{d(L - L_0)}{dt} = -k(L - L_0), \tag{3}$$

where L_0 is the center of the area of depolarization due to the inspection figure, and k is a constant.

We refer our conceptual scheme to observable variables by the assumption that the displacement, D, of a setting, observed experimentally, is directly proportional, by way of a constant, m, to the distance, $(L–L_0)$, i.e.,

$$D = m(L - L_0). \tag{4}$$

From equations [3] and [4] it follows that

$$\frac{1}{m}\cdot\frac{dD}{dt} = -\frac{kD}{m}; \tag{5}$$

and on cancelling the m's on both sides and rearranging we have

$$\frac{dD}{D} = -kdt. \tag{6}$$

On integrating, equation [6] becomes

$$\log_e D = -kt + C \tag{7}$$

or its equivalent,

$$D = ce^{-kt}$$

where C is the constant of integration, and c its antilogarithm. We may evaluate c by noting that when t is zero, c is equal to D_0, the value of D at the start of the process. Thus,

$$D = D_0e^{-kt}; \tag{8}$$

---

[3] The reader should not consider the following account as a serious statement of theory; it is presented to exemplify the use of transformations in theory testing; no claim is made as to its precise value for perception. For theoretical background, see Köhler and Wallach (*Proc. Amer. Philos. Soc.*, 1944, 88, 269–357.)

and our theory states that the magnitude of the displacement of the equality setting is a negative exponential function of the time interval between the end of the inspection period and moment at which the setting is made. By taking the more convenient common logarithms of [8] we observe that

$$\log_{10} D = \log_{10} D_0 - 0.4343kt. \qquad [9]$$

Since equation [9] is linear in form, we may test the agreement between data and theory by plotting $\log_{10} D$ against t and observing whether the experimental points may be represented by a straight line. The slope of such a plot is 0.4343 k and the intercept is $\log_{10} D_0$. Thus a logarithmic transformation of the dependent variable may be used conveniently in the present analysis.

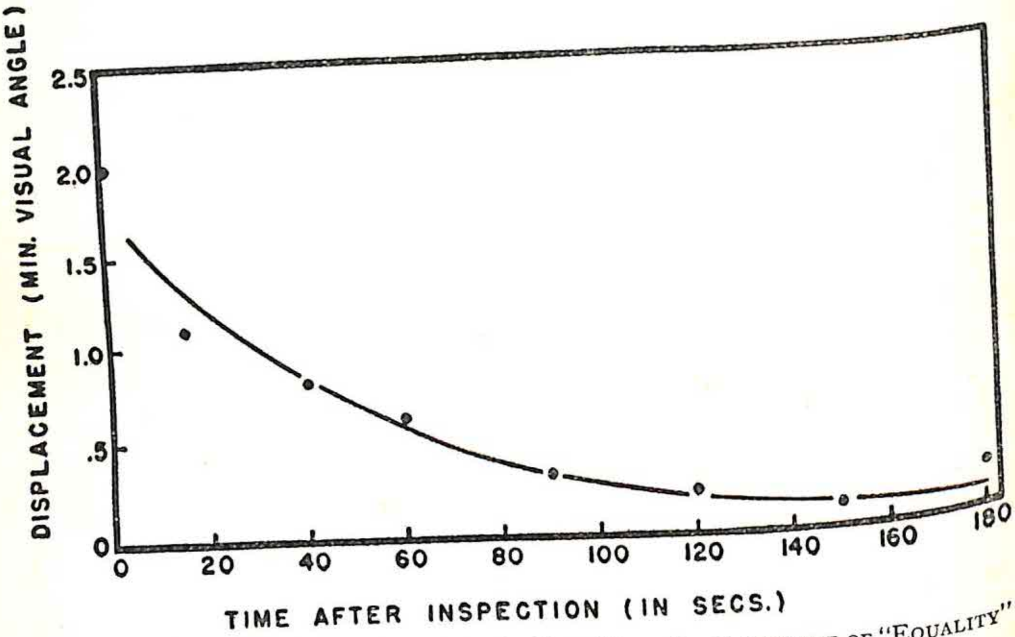


FIG. 1. HAMMER'S DATA [TABLE I IN (24)] ON MEAN DISPLACEMENT OF "EQUALITY" SETTINGS AS A FUNCTION OF TIME AFTER INSPECTION PERIOD.

Hammer (24) has investigated the magnitude of the displacement of setting as a function of time after the inspection period, and Fig. 1 shows a plot of some of her data (her Table I, data for Group Mean, Right Displacements). Fig. 2 shows a semi-logarithmic plot of these same data. With the exception of the point at 180 seconds, the relation in Fig. 2 may be represented by a straight line. The deviant point at 180 seconds is not far removed from the line of fit (calculated from the straight line plot) through the original data of Fig. 1 and lies well within the limit of variability found by Hammer (about 1.4 minutes).

3. Of course, not all problems are as simple as the examples given, and in many cases a major difficulty in testing a quantitative theoretical statement arises from attempts to determine a transformation appropriate to the solution of certain constants. An example of increased complexity in the handling of

transformations is found in a paper by Smith (43) on visual intensity discrimination. The problem was one of testing the agreement between experimental data on human intensity discrimination and Hecht's theoretical account (26).

Hecht's equation relating the adapting intensity, I, and the just discriminable difference in intensity, ΔI, is

$$\frac{\Delta I}{I} = \frac{c}{a^2 k_2}\left[1 + \frac{1}{(KI)^{1/2}}\right]^2, \tag{10}$$

where c, a, k_2, and K are constants. Smith proceeded in the following way. On taking the square root of both sides of [10], the equation may be written

$$\sqrt{\frac{\Delta I}{I}} = \sqrt{\frac{c}{a^2 k_2}} + \sqrt{\frac{c}{a^2 k_2}}\left[\frac{1}{\sqrt{K}} \cdot \frac{1}{\sqrt{I}}\right],$$

or, in its equivalent form,

$$\sqrt{\frac{\Delta I}{I}} = C + \frac{C'}{\sqrt{I}},$$

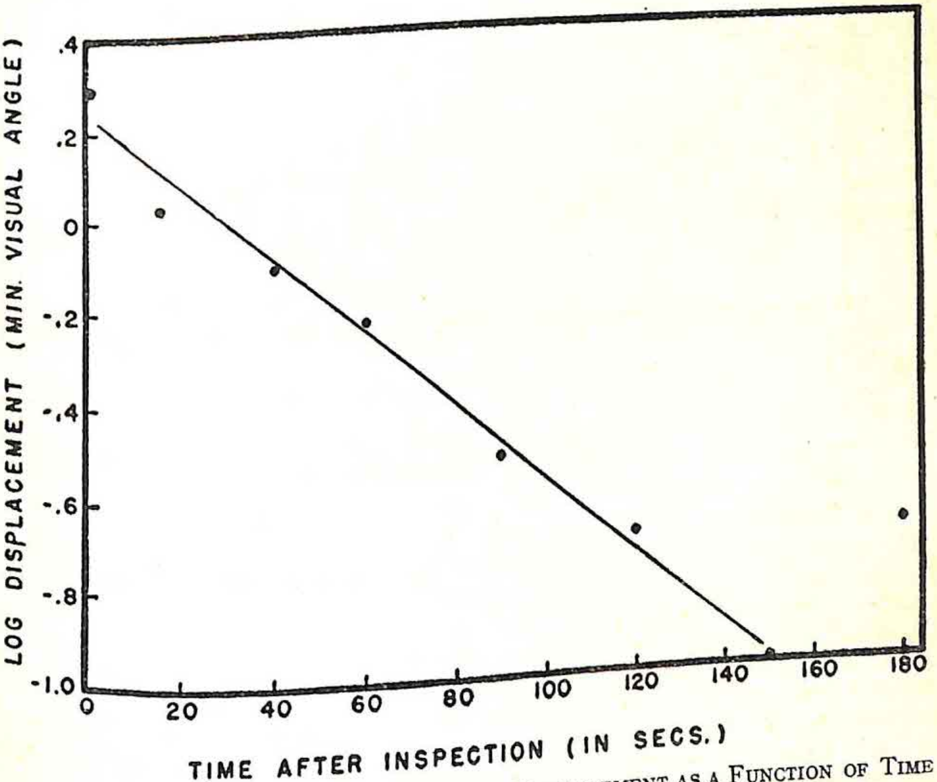


FIG. 2. COMMON LOGARITHM OF THE MEAN DISPLACEMENT AS A FUNCTION OF TIME AFTER INSPECTION PERIOD.

Data from Figure 1.

where

$$C = \sqrt{\frac{c}{a^2 k_2}} \quad \text{and} \quad C' = \frac{C}{\sqrt{K}} \cdot$$

Multiplying both sides of the latter equation by the square root of I leads to the expression

$$\sqrt{\Delta I} = C\sqrt{I} + C'. \tag{11}$$

By this development, it is shown that theory demands that a plot of the square root of ΔI against the square root of I should give a straight line of slope C and intercept constant C'. Since the range of intensities is large and the experimental points are not equally spaced in the interval, a more conveinent form of [11] is

$$\log_{10} (\sqrt{\Delta I} - C') = \log_{10} \sqrt{I} + \log_{10} C, \tag{12}$$

when C' is determined from [11] after evaluating C from the asymptote of a plot of $\Delta I/I$ against I. (C is the square root of the asymptote of such a plot and may be estimated by visual trial.) Equation [12] represents a linear function with a slope of 1.0. It is the form used by Smith and involves a number of elementary transformations, including subtracting a constant and taking the logarithms and square roots of numbers. Fig. 3 shows the agreement between Smith's data and Hecht's theory.

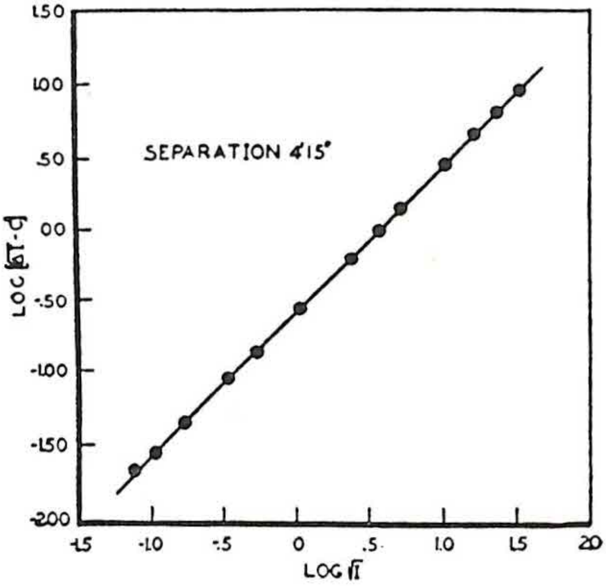


FIG. 3. LOG $\sqrt{\Delta I}$-C) PLOTTED AGAINST LOG $\sqrt{I}$.
In accord with theory this curve has a slope of unity. Fig. 8 in Smith (43).

TRANSFORMATIONS TO AID DESCRIPTION

In some cases transformations may be applied for the sole purpose of finding an approximate descriptive expression for the relations between variables. A problem may, for example, demand a quantitative statement of the approximate relationship between two variables for purposes of designing equipment, controlling certain forms of behavior, and the like. As opposed to the situation discussed in the preceding section, such a quantitative expression would, in general, have no theoretical basis, but would describe the data within the limits set by the consistency of the data and considerations of simplicity involved in the selection of an expression.

One method of arriving at an empirical formula for any particular set of data is to apply various transformations to one or both of the variables until a set of manipulations is found which yields a straight line. The transformations finally selected as successful in this respect may provide an empirical statement of the way in which the original variables are related. It is important to note, of course, that these computational procedures alone imply no theory, and the resulting descriptive expression in no way constitutes a theory of the data under consideration.

1. An experiment on the relation of area of a visual stimulus to the absolute threshold provided the data represented in Curve A, Fig. 4, where values of intensity of the visual stimulus (at threshold) are plotted against values of area, selected from a much wider range (Graham, Brown, and Mote, **20**). In our search for an empirical expression to describe these data we may systematically investigate the simpler and more common transformations with respect to their success in giving linearly related variables. In the present instance the curve appears hyperbolic, and a reciprocal transformation of one of the variables (area) results in a straight line relationship with an intercept value of zero, as shown in Curve B, Fig. 4. This finding, of course, implies that the product of the original variables is a constant, that is

$$A \cdot I = k, \tag{13}$$

where A is the area of the stimulus, I is the intensity at threshold, and k is a constant. This formula expresses Ricco's law which holds, approximately, for small area in the fovea and periphery; it has been shown to be seriously in error over large ranges of stimulus area (Graham, Brown, and Mote, **20**; Wald, **46**). Equation [13] in no way constitutes an "explanation" or a theory of area-intensity relations; the operations we have performed give it no rational basis. Within a restricted range of data it may provide an approximate correction factor for comparing thresholds of differently sized areas or adequately serve some other practical purpose. Graham, Brown, and Mote have shown that, on rational grounds, an entirely different sort of equation fits the data over a much larger range of stimulus area.

2. Another illustration of the procedure under consideration may be taken

from the data of Schoenfeld, Antonitis and Bersh (41) which are presented in Fig. 5, where the median number of responses (bar-pressing responses in a modified Skinner apparatus) in a one hour test period are shown for six successive days. The animals had no history of reinforcement in this situation and were not reinforced during test sessions. The data therefore represent the "operant rate of responding" on successive days. We guess that the data are exponential

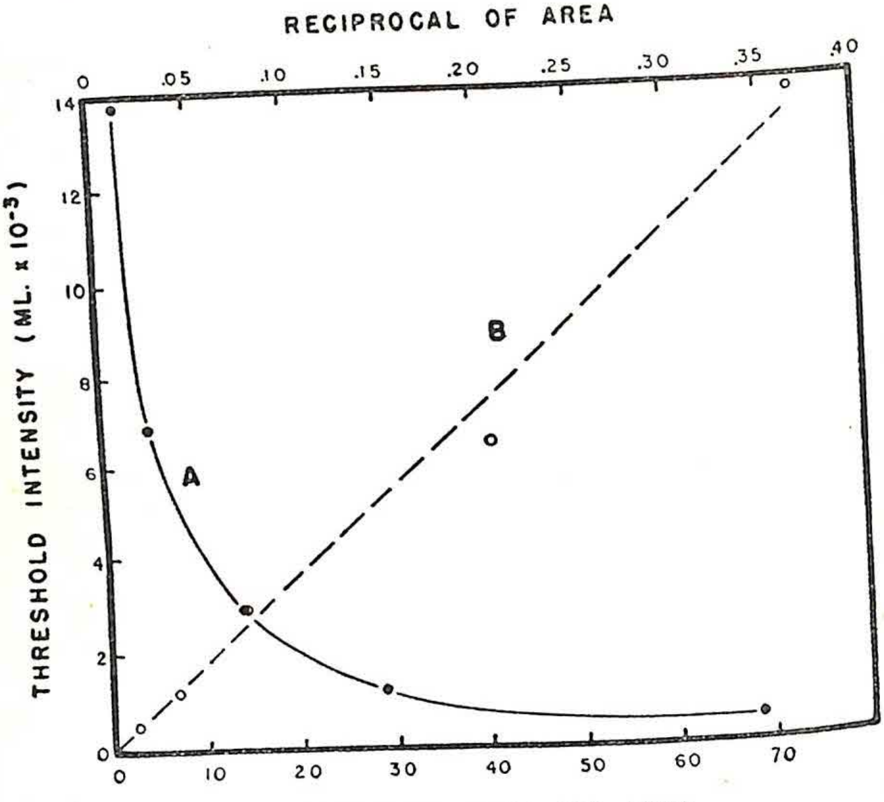


FIG. 4. THRESHOLD INTENSITY PLOTTED AGAINST AREA (CURVE A) AND RECIPROCAL OF AREA (CURVE B).

Data from Table II in Graham, Brown, and Mote (20).

in form, and so we take the logarithms of the ordinate values as a starting point in seeking an empirical description of the results. Since the curve in Fig. 5 seems to approach an asymptote which is not zero, a constant (the asymptotic value) must be subtracted from the number of responses before the logarithms are taken. If we take this asymptotic value to be 9.0, subtract it from each of the median values plotted in Fig. 5 and take the logarithm of the difference, we obtain the results shown in Fig. 6. The data deviate little from linearity which indicates that the original data may be approximately described by an exponential equation. Obviously the equation has significance only for integral

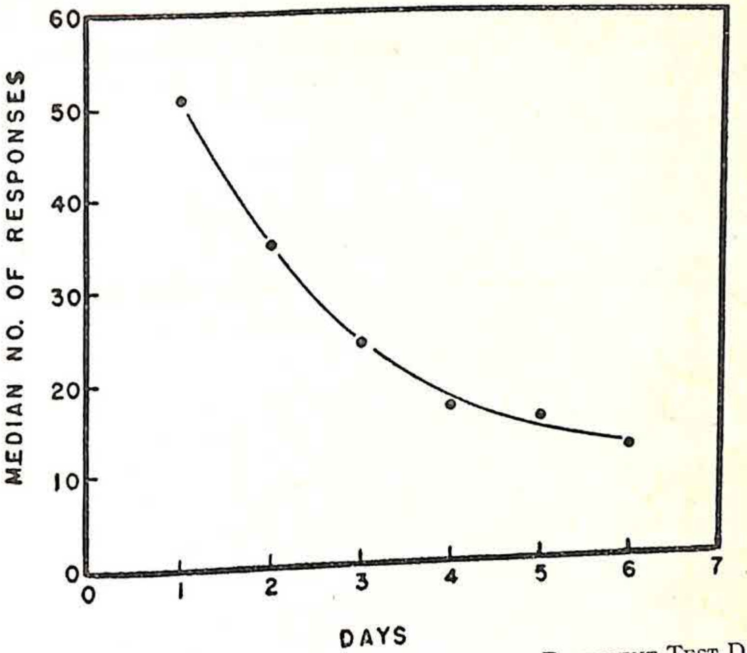


FIG. 5. MEDIAN NUMBER OF RESPONSES IN AN HOUR ON DIFFERENT TEST DAYS. Data from Schoenfeld, Antonitis, and Bersh (41).

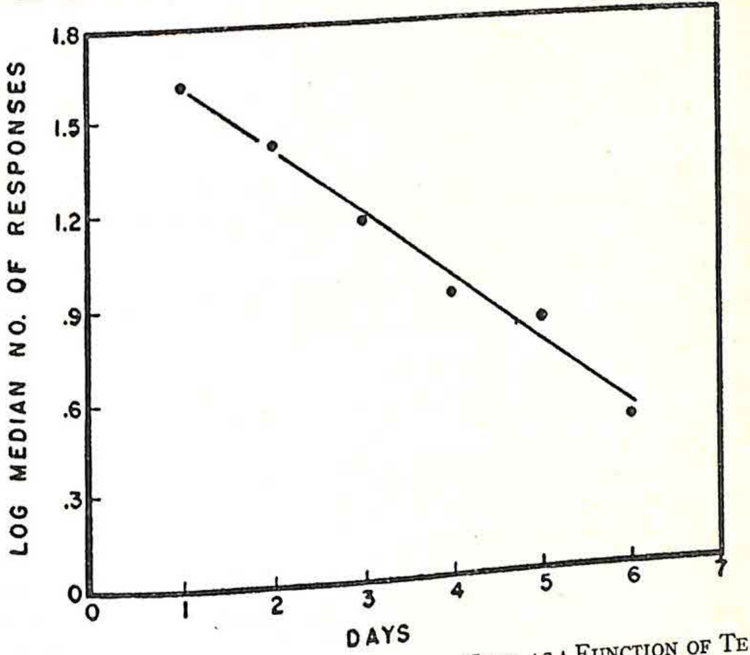


FIG. 6. LOG MEDIAN NUMBER OF RESPONSES IN AN HOUR AS A FUNCTION OF TEST DAY. Data from Fig. 5.

abscissa values. Theoretically, an infinity of curves might also represent the data. Possibly an hyperbolic expression would serve as well as the exponential and would be no more complex.

TRANSFORMATION FOR STATISTICAL TEST

A third situation in which one may encounter a need for transformations is in the case where adjustments are made to permit the use of a particular statistical test without violating the assumptions involved in the efficient use of the test. This class of problem has been of considerable concern to mathematical statisticians and applications of their theoretical work may be found in the scientific literature (Beall, **4**; Eisenhart, Hastay, and Wallis, **13**; Williams, **47**).

Univariate distributions. Consider a simple example of an experiment in which we measure the reaction time to the onset of a specified visual stimulus under several conditions. Suppose our experiment has yielded several sets of data, distributions of number of reaction times of various magnitudes. In such an example we are dealing with a univariate distribution, i.e., a frequency distribution with respect to a single variable. Since experiments on reaction time (e. g., Jenkins, **29**) have shown that, under many conditions, the frequency distributions of reaction time are not normal but positively skewed, problems may develop in the application of many statistical tests. If marked skewness exists, a direct application of "normal curve statistics" would undoubtedly lead to error because of the lack of correspondence between our data and the conditions imposed by the test.[4] However, several procedures for avoiding this error are open to us: we may (1) use a statistical test applicable to our type of distribution if such a test exists, (2) use a test which makes no assumptions about the population from which our sample is drawn, or (3) transform our scores into a normal distribution so that we may use one of the many tests available for this distribution. It is with the latter procedure that we are here concerned.[5]

[4] The degree of error, of course, is the important thing in any specific situation. If the error is not large the experimentalist will be willing to ignore it in the interest of simplicity of calculation. There is some evidence [see Smith and Duncan (**42**) and Cochran (**7**) for discussion and references] that minor deviations from normality in fairly large samples do not seriously impair the sensitivity of certain statistical tests, but unrestricted generalization of belief in this state of affairs to all cases of non-normality, especially in small samples, is certainly to be avoided. Evidence is clear that, in small samples, serious error is introduced by ignoring the form of the population distribution [Rider (**39**), Festinger (**16**); again see Smith and Duncan (**42**) and Cochran (**8**) for discussion and references].

[5] It is not the purpose of this paper to discuss all of the methods of handling data from populations having non-normal or "unassumable" characteristics, but some ori-

The general problem that confronts us is one of transforming one set of scores having some specified frequency distribution, $y=f(x)$, into another set having a distribution, $y=F(z)$. It is required that we find some function $z=g(x)$ such that when transformations are made from x to z, z is distributed in a manner specified by $y=F(z)$.

Although conversions of a set of scores into any specified frequency function is possible, the most frequently encountered problem is one of making the data approximate the normal curve. The interest in having the data approximate the normal curve arises mainly from the greater amount of information available on the sampling characteristics of the parameters of this function; if the data can be put into the appropriate form, larger resources exist for testing statistical hypotheses.

At least two possibilities are open to us in our search for an appropriate transformation. The first presents us with a solution of the general problem to an approximation which can be made as exact as we please. Several solutions of this type are available and involve an approximation to the desired function by using a power polynomial or series function [cf. Baker (**1**), Fry (**19**), Kendall (**30**)]. These solutions are general and may be used on any continuous distribution. The second possibility available to us is to see whether one of several convenient transformations, such as taking the logarithms or reciprocals of the scores, gives an appropriate and sufficiently precise solution. Logarithms or reciprocals provide an exact solution in those cases in which $z=g(x)$ is a logarithmic or reciprocal function; practically they have been found to

---

entation with respect to the use of transformations may be gained if some of the other methods of handling data are mentioned. Although the emphasis on sampling distributions of parameters of non-normal populations has not approached that given to normal curve functions, work on non-normal distributions [e.g., Festinger (**16**), Paulson (**35**)] may eventually decrease the advantages of recoding scores to test statistical hypotheses by providing analogous tests for non-normal functions. In the present case of a skewed distribution of reaction time scores we might test for the differences between means in two such distributions by using a test suggested by Festinger (**16**). Unfortunately such techniques are not numerous.

The selection of a statistical test which makes no assumptions about the population distribution will probably become an increasingly important procedure to research workers. If we are willing to sacrifice some of the information present in our scores by dealing with ranks or orders of scores, statistical procedures are available for testing many hypotheses, e.g., that population correlation is zero [Hotelling and Pabst (**28**); Pitman (**37**)] and that two samples came from the same population [Wald and Wolfowitz (**45**); Pitman (**36**); Mann and Whitney (**32**); Festinger (**17**); Dixon (**12**)]. In cases where normal curve statistics are appropriate they are more powerful than non-parametric tests, since the latter do not use all the available information; but the generality of statistical tests of the latter type presents a definite advantage to the research worker.

provide sufficiently accurate approximations in a large number of cases encountered experimentally.

If we were interested in approximately "normalizing" our reaction time data, some convenient transformation may serve as a first step. Jenkins (**29**) has shown that the logarithms of reaction time scores are approximately normally distributed, and this transformation may be our first choice. If we had two distributions on a single dimension (such as two sets of reaction time scores) and were interested in significance of the differences between means, such a transformation would be an appropriate step preliminary to the application of a *t*-test or some similar test of significance.

Bivariate distributions. In the field of statistical analysis one frequently encounters the variate transformation in situations which involve two variables rather than one.

Transformations, as applied to bivariate distributions, become an important tool of analysis in two cases: (a) in the fitting of theoretical or empirical curves to data, and (b) in the application of tests of significance involving an "analysis of variance." Most applications of "analysis of variance" techniques and most fitting of curves by the method of "least squares" involve the assumptions of normality and of homogeneity of variance in the various columns.[6] If the heterogeneity of variance exists and results from a correlation between mean values and standard deviation values, it is frequently possible to find a simple transformation by which variability may be equalized. Examples of useful transformations in the application of "analysis of variance" tests have been extensively discussed by Bartlett (**2**, **3**), Cochran (**6**), Curtiss (**12**), and others. In the present account we consider a few examples taken from the field of psychology.

1. Many groups of data in psychology show a variability which is approximately proportional to the mean. For example, the measurement of differential thresholds in sensory psychology provides data demonstrating a proportionality between the mean and standard deviation over a considerable range of the controlling variable, intensity. If we attempt to analyze the effect of certain variables on mean differential threshold by using "analysis of variance" techniques we require some method for equating the variances in the several classes. In such a situation it may be shown that the variability is *roughly* equalized by a logarithmic transformation.

Consider an example in a visual intensity discrimination experiment of finding a transformation such that the variability of ΔI is constant for various values of ΔI. Crozier (**10**) has shown that the standard deviation of ΔI is approximately proportional to ΔI, that is,

---

[6] Exceptions occur in "analysis of variance" tests of the non-parametric type [Pitman (**38**)] and in "least squares" solutions utilizing a weighting term, as we shall see below.

$$\sigma_{\Delta I} = k\Delta I, \qquad [14]$$

where k is a constant. We require a transformation to y, where $y = f(\Delta I)$, such that the standard deviation of y is constant for all mean values of y. The variance of a function of a variable can be represented *approximately*[7] (**30**) by the equation

$$\sigma_z^2 = \sigma_x^2(dz/dx)^2, \qquad [15]$$

where x is the variable and $z = f(x)$. From [14] and [15] it follows that

$$\sigma_y = k\Delta I[dy/d(\Delta I)], \qquad 16]$$

where $y = f(\Delta I)$. If we let σ_y equal a constant, a, and transpose, we obtain

$$dy/d(\Delta I) = a/k\Delta I. \qquad [17]$$

The function, $y = f(\Delta I)$, which has the latter derivative is $(a/k)\log_e \Delta I$; and so logarithms of ΔI are the appropriate numbers to use in this type of experiment when homogeneity of variance is presupposed.

2. Sensory psychology is not the only area from which examples may be drawn; many studies in conditioning and learning, for example, present similar problems. As an example taken from the field of conditioning we may examine the data of Zeaman (**48**) on the changes in latency of a running response during acquisition.[8] Zeaman used a straight runway of the Graham-Gagné type and included in his measures the time interval between the opening of the door to the runway and the entry of the animal onto the runway. An examination of the raw data shows (1) that the distribution of latency scores for a group of animals on any particular trial is markedly skewed and (2) that the variability of each distribution varies with the mean value.

Assume that we are interested in arriving at a "best fitting" line to represent the data or in testing the hypothesis that there are no significant changes in average latency from trial to trial. Most methods of satisfying these interests require that the distribution of scores for each trial be normal and that the variability be constant from trial to trial. This is not the case with the latency scores, so we seek some transformation that will yield scores satisfying these conditions. We hope that the transformation that stabilizes the variability in the various trials also normalizes the distributions for each trial. In view of the success of the logarithmic transformation in approximately "normalizing" some reaction time data, we may begin our analysis by taking the logarithms of the latency measures. To test the success of this device in performing the desired function let us cumulate the percentages of cases falling above specified log latency values and plot this cumulated percentage on probability paper with log latency on the abscissa. The cumulative plot of a normal distribution should give a straight line whose slope is proportional to the standard deviation. Such a plot of Zeaman's data is shown in Fig. 7. The data for the odd numbered trials are shown. The approximate normality of the distributions of scores for each trial is obvious from the fact that the points may be approximately represented by straight lines. The homogeneity of variance is demonstrated by the fact that the lines through the points for various trials are nearly parallel. The displacement of the lines along the abscissa shows the changes in the average log latency during the course of acquisition. This simple transformation has given us a function in Y (log latency) and X (trials) which has the characteristics that, for changes in X, we obtain changes in the mean value of Y but no appreciable changes in the variability of Y or in the form of the distribution of Y values. This result confirms a much earlier report by Graham and Gagné (21) on the advantages of using the logarithms of the latency measures in a similar situation.

---

[7] Equation [10] and the development that follows involve several severe approximations; they are presented for their heuristic value to those not familiar with the argument. For some difficulties with this development and a more rigorous account of the transformation discussed here (and several others) see Curtiss (**11**).

In addition to the approximations involved it is important to note that the operations outlined here provide no guarantee of normality even though they purport to stabilize variability. In those cases where we are dealing with particular forms of skewness, such as Poisson distribution, logarithmico-normal distribution, etc., the transformation which stabilizes variability will also normalize the data. Forms of variance heterogeneity are imaginable, however, in which this would not be the case.

[8] The author wishes to express his gratidue to Dr. David Zeaman for making the data available but assumes full responsibility for any analyses or conclusions that result from their consideration here.

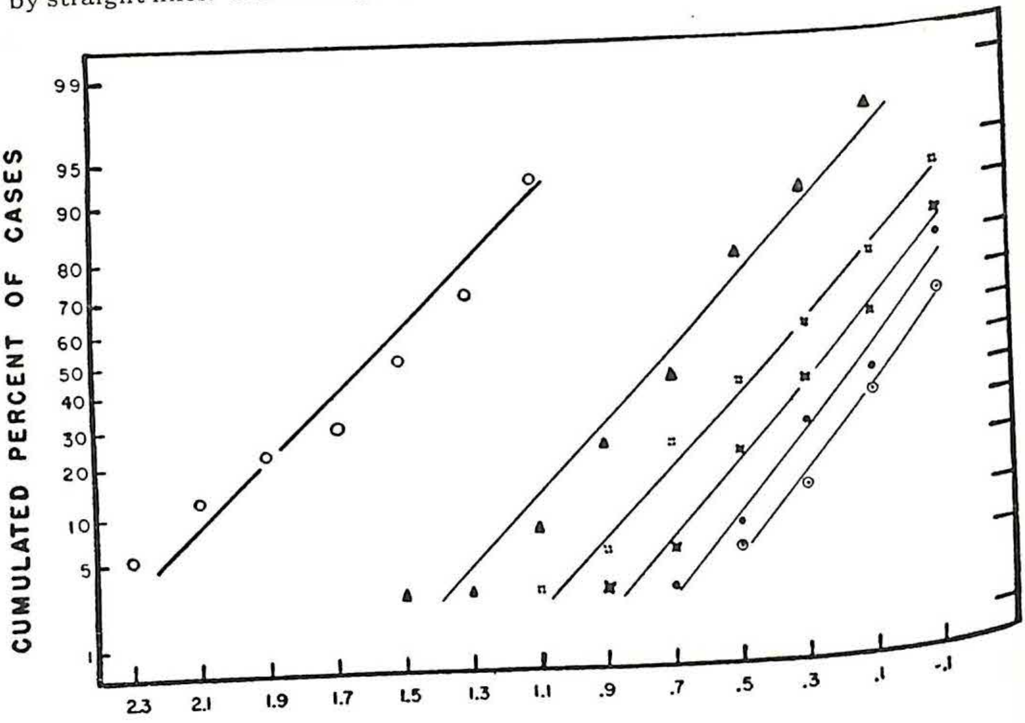


FIG. 7. ZEAMAN'S DATA (48) ON ODD NUMBERED TRIALS (1 TO 11), WITH CUMULATED PER CENT OF CASES (PER CENT OF CASES FALLING ABOVE THE ABSCISSA VALUE) PLOTTED ON THE ORDINATE AND THE COMMON LOGARITHM OF THE LATENCY, IN SECONDS, ON THE ABSCISSA.

The lines, from left to right, represent trials 1, 3, 5, 7, 9, and 11, respectively.

As has been frequently emphasized in the statistical literature, if a group of scores is normally distributed along some dimension, then scores along any non-linear function of this dimension will not be normally distributed. In the present instance the fact that the logarithms of the latency scores are distributed normally means that the latency scores themselves are not normally distributed. A plot, on probability paper, of the cumulative frequency of the latency scores against the magnitude of the latency shows a markedly curved line which shows deviation from normality, and the degree of curvature varies with mean value or position on the probability grid. The existence of such information, of course, argues against the application of normal curve statistics to the raw scores.

3. The examples considered so far by no means exhaust the areas in which transformations are, or should be, an important step in analyzing experimental data. Space permits only brief mention of some other examples. An important paper by Morgan (33) shows the advantage to be gained by transforming data obtained in experiments on food hoarding in rats. Morgan shows that the logarithms of the number of pellets horded are distributed more normally than are the numbers themselves and that variance under the various conditions is more nearly equal with transformed scores.

Haggard (22) and Haggard and Garner (23) have considered the problem of the transformation of experimental data in some detail in a discussion of the analysis of the results on the galvanic skin response. Haggard and Garner show that the variability of the GSR measures is approximately proportional to the mean value of the GSR, a consideration which, as we have seen above, leads to a logarithmic transformation if we seek scores showing equal variability. In order to scale the GSR measures, Haggard and Garner employ an additional computation which need not concern us here.[9]

---

[9] Exactly how the transformation finally selected equalizes variability (in addition to equalizing the response magnitude to a constant stimulus) is not clear. Haggard and Garner present data on (1) the magnitude of the galvanic skin response as a function of the resting resistance and (2) the variability of the GSR. The data on the magnitudes of the response as a function of resting resistance is represented by an equation of the form

$$\Delta R = ae^{bR}, \qquad [a]$$

where ΔR is the change in resistance when the GSR is measured; R, the level of resistance just before the response occurs; a and b, constants; and e, the base of Napierian logarithms. From [a] it follows that

$$\log_e \Delta R = \log_e a + bR \qquad b]$$

and that

$$\frac{\log_e \Delta R - \log_e a}{R} = b. \qquad [c]$$

The left side of equation [c] indicates the transformation used. Haggard and Garner summarize the data on the variability of the response by the statement that the standard deviation of the response magnitudes is approximately proportional to the mean value. We have seen above that under such circumstances the logarithms of the ΔR values should therefore show equal variability. If we then divide these log values by a variable term, R, as dictated by equation [c], variance heterogeneity should result. Computations by Haggard and Garner indicate that this is not the case.

SOME THEORETICAL CONSIDERATIONS

Transformations required by theory. The selection of the appropriate transformation to "normalize" the data or to equalize the variability in the cases discussed so far has been on the basis of empirical considerations; we have been interested in satisfying certain conditions, as, for example, the requirement that the transformed scores be normally distributed and/or show homoscedasticity. On this basis we sought transformations which would serve our purpose. Obviously, there are many transformations which will satisfy any specified condition to any level of approximation we may designate, and our selection from among these possibilities has been on the basis of simplicity, ease of calculation, and degree of approximation desired. In certain instances, however, other considerations may enter to determine the form in which the data may be most fruitfully analyzed, and it is here that we encounter the role of scientific theory in restricting the range of selection of the appropriate dimension for analysis.

Earlier it has been shown that, where there is evidence that the magnitude of variability is roughly proportional to the mean value, we may approximate homogeneity of variance by making a logarithmic transformation of the data. Frequently we are led to make such a transformation on the more indirect basis of a successful theoretical formulation in the field with which our experiment is concerned, even though our data may not contain conclusive evidence that variance heterogeneity would occur without such a transformation.

1. Consider the case in which changes in some dependent variable are a function not only of the changes in the independent variable but also some function of the magnitude of the dependent variable. A model for such a relationship may be found in many photochemical systems, for example, those in which the amount of decomposition is a function not only of the intensity of light but also the amount of photosensitive material available.

A general form of the kind of theoretical statement frequently encountered in photochemistry is

$$-\,dy/dt = kIf(y) \qquad [18]$$

where y is the amount of photosensitive material, I is the light intensity, k is a constant, and $f(y)$ is usually y or y^2. Rearrangement of [18] yields

$$-\,kIdt = dy/f(y),$$

from which it follows that

$$-\,kIt = \int dy/f(y). \qquad [19]$$

It may be shown (Cramer, 9) that, if variations in the left member of [19] as the activities of independent random variables then, in the limit, the integral of the right member can be represented in the same way. If we take, as an example, the simple case in which $f(y)$ is equal to y, that is, the rate of change is proportional to y, then the right side of equation [19] reduces to $\int dy/y$ or $\log y$. Thus, if variations in It follow a normal Gaussian function, we should find that the logarithms of y values would be similarly distributed, and analysis involving y must consider this possibility.

2. Although psychology does not provide many examples of transformations which are derivable from theoretical considerations, some cases may exist. Let us reconsider the food hoarding data discussed by Morgan (**33**). It will be recalled that Morgan found a logarithmic transformation to be useful in treating his data. Assume that, for any experimental condition, i.e., one representing constant drive level, illumination conditions, amount of infant feeding deprivation, etc., there exists a fixed rate of occurrence of the hoarding response and that the responses under any fixed experimental conditions are distributed randomly in time. For events distributed randomly in time, the probability (P_n) of obtaining n events in a time interval, t, is

$$P_n = \frac{(rt)^n e^{-rt}}{n!}, \tag{20}$$

where r is the rate of occurrence of the events and e is the base of Napierian logarithms. Thus, the number of events observed in intervals of fixed length (experimental test interval) will be distributed according to the Poisson law. One characteristic of the Poisson distribution is that the mean is equal to the variance, which means that any condition which changes the mean thereby changes the variability. By reasoning similar to that used in finding the transformation for the case where the standard deviation is proportional to the mean, we find that the transformation that approximately equalizes the variability of a set of Poisson distributions is the square root transformation. [Again see Curtiss (**11**) for a rigorous proof.] Thus, from the assumption that the responses are randomly distributed in time for any experimental test condition, we may employ the square roots of the number of hoarding responses as data in situations where homogeneity of variance is presupposed. The author knows of no test of this notion, but presents it as an example of the kind of consideration which may enter.

The reduction of "laws" by the use of transformations. Although the rationale of most deliberate attempts to stabilize the variability of one variable, Y, for various values of another variable, X, may be found in the application of statistical tests, the demands of statistical tests may not be the only, or even the most important, reason for seeking homogeneity of variability. One characteristic of a bivariate frequency function with normal distributions of equal variability in the various columns has not been emphasized in the literature. When the values of one variable, Y, are such that they are normally distributed with equal variability for all values of X, the function relating the two variables, Y and X, does not change in form under translation of the measure o position, i.e., under a change of the representative value employed (mean, median, mode, centiles, etc.). Although there are statistical bases for choosing among the many available measures of position, these bases may not appear compelling to the theorist if other considerations of importance enter, such as the observation that different functional relations between variables are obtained depending on the measure used. For example, the mean may be rationalized as the most appropriate measure of position *for the normal distribution* because it is the most efficient statistic, i.e., no other statistic has a smaller sampling variability. The differences in efficiency, as for example between the mean and median, however, are hardly of a magnitude to dictate the particular form of analysis if complexities of theory result. When the sets of Y values for various X values yield normal distributions of equal variability no problems of the multiplicity of functional relations arise, since the mean, median, and mode have the same value and the selection of any centile or other fixed position in the distribution merely results in a displacement of the function on the ordinate, or Y, scale by some specifiable amount. The correlation between the mean and standard deviation in cases of skewed distributions, however, guarantees heterogeneity and greatly increases the likelihood of actually finding a multiplicity of functional relations upon changing the positional measure. Thus, if our original data are skewed we could show, for a single experimental condition, as many "empirical behavioral laws" as there are measures of position for a distribution.

It seems highly unlikely that theoretical psychologists will find it fruitful to attempt to formulate the infinite set of theoretical statements necessary to handle all the "laws" that would result from any changes in the measure of position or central tendency. In the case of skewed heteroscedastic distributions it would seem much more feasible to provide a rational account of the transformation required to yield a homogeneous set of normal distributions.

An example of the difficulties encountered when dealing with non-normal distributions is available in the study by Felsinger, Gladstone, Yamaguchi, and Hull (**15**) on reaction latency. Curve A in Fig. 8 shows the cumulated distribution of reaction latencies presented by these authors plotted on probabilty paper. The distribution is obviously skewed (indicated by the marked curvature) and confirms the reports of other investigators (Graham and Gagné, **21**; Zeaman, **48**). In the light of the skewness of the data, it is not surprising that the authors find different functional relations between reaction latency and number of reinforcements if the mean or the median of the distribution of latency scores is taken as the representative value. As in the case of Zeaman's latency data in Fig. 7, we find that the logarithms of the time measures are *approximately* normally distributed as shown by the linearity of Curve B in

Fig. 8, which represents a cumulative percentage plot on probability paper with log latency values on the abscissa.[10]

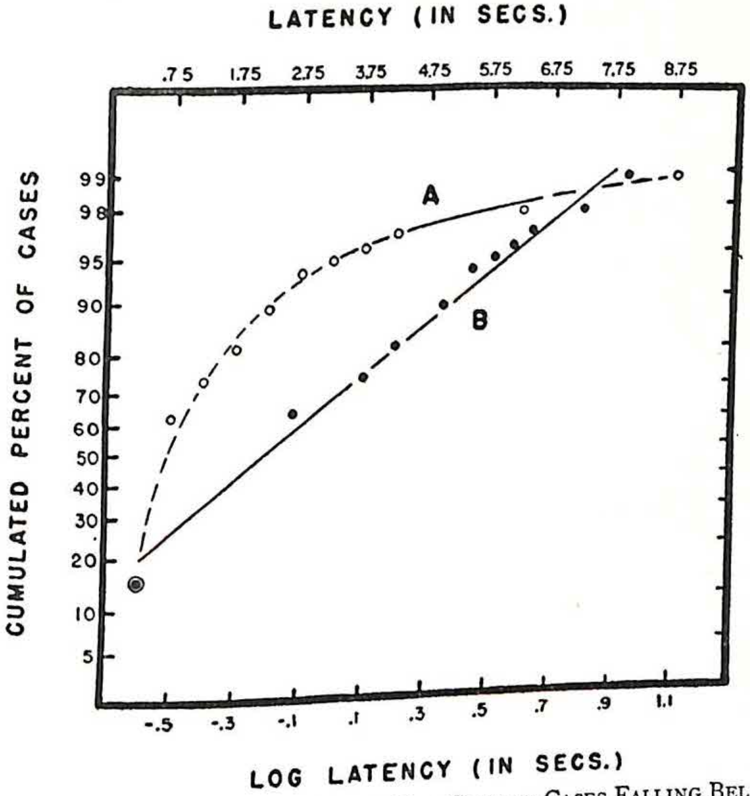


FIG. 8. CUMULATED PER CENT OF CASES (PER CENT OF CASES FALLING BELOW ABSCISSA VALUE) AS A FUNCTION OF THE LATENCY (CURVE A) AND THE LOGARITHM OF THE LATENCY (CURVE B).

Data from Felsinger, Gladstone, Yamaguchi, and Hull (15).

---

[10] If we are dealing with a logarithmico-normal distribution, the arithmetic mean would not have its usual advantage of statistical efficiency, since in this case the geometric mean shows a smaller sampling variability.

Since we have attempted to solve several problems by the use of logarithms, it is important to emphasize that the present discussion must not, under any circumstances, be interpreted as holding that the logarithmic transformation is a panacea for the experimentalist's analytical problems. The logarithmic transformation does have wide applicability, but obviously the selection of a given transformation in any situation must be made on the basis of merit.

In many of the areas from which we have drawn examples, there is evidence that other simple transformations may satisfy our requirements or that additional conversions will give closer approximations to normality. For example, there is some evidence that log reaction time scores retain some of the positive skewness of the original distributions and that a log-log transformation gives a better approximation to normality. This may also be true where latency measures are employed. In addition, deviations of the sort

SOME CONSEQUENCES OF THE USE OF TRANSFORMATIONS

Frequently we find that the selection of a transformation to simplify the form of a theoretical statement may preclude the possibility of normalizing the data or equalizing the variability. Thus, transforming our data for theoretical reasons may complicate the process of meeting the assumptions of statistical procedures.

Curve fitting. Consider the case of fitting an empirical or theoretical curve to the data of Fig. 1. Assume that we wish to fit the curve $D = D_0e^{-kt}$ to the data in Fig. 1, where t represents the time since the end of the inspection period and D represents the displacement of the setting. We have seen earlier (equations [8] and [9]) that one method of testing the applicability of such a function and, at the same time, providing an evaluation of its constants is to plot $\log_{10} D$ against t. Such a plot should give a straight line with a slope of $-.4343\ k$ and an intercept constant of $\log_{10}D_0$. If we are satisfied with an approximate indication of the relationship between the variables and a subjective estimate of the degree of fit, fitting the line by visual inspection satisfies our needs. For the purposes of the present discussion, however, let us consider a more "exact" method of evaluating the constants involved and note some of the problems that arise from such a consideration.

Our first problem arises in the selection of a statistical solution of the "best fitting" line. Several techniques for obtaining this "best fit" are available, but we shall arbitrarily select one for discussion, the method of least squares. This method serves as a good example and has the advantage that it is widely employed and has been extensively developed. The method is characterized by the assumption that the "best fitting" line is that line about which the weighted squared deviations of data from the line give a minimum sum.

---

shown in the first trial in Fig. 7 may prove to be replicable. For the special cases of the logarithmico-normal distribution where the lower limit of the scores is not zero, the subtraction of a constant (which is the lower limit of the scores) from each score before the logarithms of the numbers are computed may be necessary to insure normality. This consideration arises from an inspection of the parameters of the generalized form of logarithmico-normal frequency function, which is (Cramer, **9**),

$$\frac{1}{\sigma(x-a)\sqrt{2\pi}}\,e^{-[\log(x-a)\,-\,m]^2/2\sigma^2} \qquad \text{[d]}$$

where σ is the standard deviation, and m, the mean of the log distribution; x, a score; a, a constant; and e, the base of Napierian logarithms. The adequacy of the simple logarithmic transformation in any of these cases will depend upon the degree of approximation required as well as the particular characteristics of the data.

In addition to this matter of emphasis, it is to be noted that many useful transformations, for example the arcsin transformation (Zubin, **49**; Eisenhart, Hastay, and Wallis, **13**), have not even been mentioned because of limitations of space.

The problem of the appropriate weighting of points is one of the important considerations in curve fitting, although it is one which arises only implicitly in most applications of the least squares method. We know that the combination of several distributions of raw scores into a single distribution effectively weights the separate distributions in proportion to their variability; that is, although we ostensibly weight each distribution equally by merely adding the scores to obtain the final distribution, implicit weighting arises from differences in variability in our original distributions. We shall see that this factor plays an important role in the solution of best fitting lines in the cases of original and transformed data, since the usual normal equations

$$\begin{aligned} aN + b\Sigma x &= \Sigma y \\ a\Sigma x + b\Sigma x^2 &= \Sigma xy \end{aligned} \qquad [21]$$

make no provision for differential weighting; that is, they assume homoscedasticity or equal variability along the fitted line. The use of such equations in fitting a line to the logarithms of the scores yields a line around which the sum of the squares of the logarithms of the scores is a minimum. If the variability of Y is equal on a logarithmic scale, i.e., if log Y values show equal variability for all values of X, each point on the curve is equally weighted. But if the Y values themselves show equal variability for various values of X, using equation [21] differentially weights each point as a function of its Y value. Thus the solutions of a and b obtained by fitting a straight line to the transformed data by the least squares formula [21] differs from the solutions obtained by fitting a curve directly to the raw data on the basis of the least squares criterion. In the present case this means that the numerical values of D_0 and k obtained from the data of Fig. 1 depend upon how we fit the exponential curve: directly to the data, or on the basis of a straight-line plot of $\log_{10} D$ against t, using equation [21].

It seems clear that the choice between the constants obtained in the one case by fitting a straight line to the transformed data by equation [21] and in the other case by direct application of least squares to the exponential form depends upon theoretical and empirical considerations; these considerations must be specifiable if the solution of the constants in the function is to have precise significance. If the least squares formula without a "weighting" term is applicable to the untransformed scores, it is not applicable to the transformed scores and vice versa.

It seems imperative that considerations of the sort just outlined be seriously entertained by workers for whom the exact numerical value of constants is important. Variability factors as well as normality factors are too important for the solution of these parameter values to expect that we can efficiently estimate population parameters and give any interpretable statement about confidence intervals without paying specific attention to these and related topics.

The generality of our "least squares" procedure is increased if we employ formulae which incorporate a term for weighting each point. The required formulae are given by the equations

$$\begin{aligned} a\Sigma w + b\Sigma wx &= \Sigma wy \\ a\Sigma wx + b\Sigma wx^2 &= \Sigma wxy \end{aligned} \qquad [22]$$

where w represents the weight and is defined as the reciprocal of the variance. The similarities of formulae [21] and [22] are obvious. In addition to the usual computations, however, equation [22] requires that we have a method for determining the variability of transformed scores when we are given the variability of the original scores.

We have seen that an approximate solution for the variance of a transformed distribution in terms of the variance of the original distribution is given by equation [15]

$$\sigma_z^2 = \sigma_x^2(dz/dx)^2,$$

where z is the transformed score; x, the original score; and σ^2, the variance of z or x scores. Consider a logarithmic transformation. If the variability is small compared with the mean, then the log of the mean of x is approximately equal to the mean of log x, in which case

$$\sigma_z^2 = (1/\bar{x})^2\sigma_x^2. \qquad [23]$$

Equation [23] allows us not only to determine weighting terms but it also allows us to evaluate the error in curve fitting introduced if account is not taken of implicit changes in weighting.

The combination of the weighted least squares formula and the general solution for a set of transformed scores greatly increases the generality of the method of curve fitting and permits other than statistical considerations to dictate when and what transformations shall be made. In particular, convenience or simplicity of the form of our theoretical statements may now play a larger part in determining the analysis of agreement between data and theory.

Correlation. Discussion of the least squares method of curve fitting leads to another example of the bivariate distribution in which similar assumptions are made and similar consequences figure. The example is the product-moment correlation. Although a formula for the correlation coefficient may be developed without assuming a normal bivariate distribution, it must be emphasized that such a proof is not a sufficient condition for the extension to all cases of the large superstructure of correlation theory which is associated with the correlation term in the normal bivariate frequency distribution, particularly that portion of correlation theory having to do with sampling, significance tests, etc. It is true that one may compute a correlation coefficient for any group of paired numbers, but the interpretation to be placed on the result of this computation is not always clear. Statements about the sampling characteristics of the correlation coefficient, inferences relating to population parameters and differences between parameters, are, for the most part, limited to normal bivariate distributions.

Our discussion points to a few factors which may be expected to limit the application of correlation techniques to non-normal populations. Two obvious changes are introduced in the normal bivariate distribution when non-linear transformations are made in one of the variables. The first of these is a change in the regression line; if the raw scores are linearly related, then the transformation of one of the variables will change this relationship. The magnitude of this change will obviously depend on the transformation made. The second change introduced is also of a type mentioned earlier, that of a change in variability of the distributions of the transformed variable. Both of these factors guarantee that at least one of the forms of bivariate distribution (the one involving transformed scores or the one with the raw scores) is not normal, and different values for the correlation coefficient will be obtained by performing similar computations on the two forms.

1. The importance of the form of the bivariate frequency distribution for product-moment correlation between variables is considered in papers by Ezekiel (**14**) and Stephan (**44**). In discussing the importance of linearity of regression in the product-moment correlation, Ezekiel points out that the logarithmic transformation of one variable frequently converts a curvilinear regression to approximate linearity and that product-moment correlations are increased in such cases. In one example, however, Ezekiel showed that such a transformation not only did not increase, but actually decreased, the correlation. The example was one in which the per cent protein in grain was correlated with the per cent vitreous kernels. As pointed out by Stephan the decrease in correlation is not surprising since the logarithmic transformation was not appropriate in this case and merely served to increase the curvature of the line relating the variables. Stephan showed, however, that by transforming one dimension to per cent non-vitreous kernels (rather than per cent vitreous) and taking the logarithms of these values, the regression became approximately linear and the variance around the straight line fitted by least squares was considerably reduced. The correlation between the original variables was .73 and the correlation after transformation was −.98. Thus, with the same experimental data and without changing the rank order of the scores, the presumed "amount of variance accounted for" was almost doubled.

2. Another example of a case in which the interpretation of experimental data would have been markedly different if an appropriate transformation had not been employed is an experiment by Mueller and Richmond (**34**) on the comparative reliabilities of several methods of measuring visual acuity. The original data, numbers recorded by the experimenter for each subject, were the denominators of the fraction used to classify each size of figure in commercial tests of visual acuity. The numerator for all such fractions is a constant and typically 20. The fraction has the dimensions of a visual acuity and has considerable empirical and theoretical status as a variable in vision. These considerations and the facts that the distributions of the fractional scores for the several tests were approximately normal (very slightly positively skewed) and the regressions between the tests approximately linear, left little doubt as to the appropriate dimensions for the reliability analysis. When the fractional scores were used in computing the test-retest correlations (product-moment), the results were as given in Table I, column *a*. An interest in the importance

TABLE I

Test-Retest Reliabilities for Six Tests of Acuity.

Column (*a*) shows the results when the acuity scores are used; column (*b*) shown the results when the reciprocals of the acuity scores are used.

Test	*Reliability*	
	a (acuity scores)	*b (reciprocal of acuity scores)*
A	.81	.81
B	.75	.79
C	.69	.86
D	.68	.76
E	.61	.81
F	.56	.65

of the form of the bivariate distribution, however, prompted an analysis of the test-retest correlations using the originally recorded scores, rather than the fraction scores. The correlations in the case of the original scores are presented in Table I, column *b*. Tests for the significance of the differences (using Fisher's *z* transformation) in Table I, column *a*, leads to the conclusion that test A has a higher reliability than tests C and D, at between the 1 per cent and 5 per cent levels, and higher than tests E and F at less than the 1 per cent level. The same significance tests applied to the data of Table I, column *b*, lead to the conclusion that test C has a higher reliability than D at between the 1 per cent and 5 per cent levels and higher than F at better than the 1 per cent level. Obviously, such a result constitutes a dilemma only if we ignore the importance of the frequency characteristics of the data to be analyzed. Actually, we have reasons to doubt the applicability of our test of significance in the case of the untransformed scores. The point to be emphasized is that test-retest correlations on the recorded data would have led to conclusions incompatible with the results of a more appropriate analysis.

Summary

This paper considers several circumstances which may call for the numerical transformation of data: (1) testing theory, (2) finding empirical descriptive expressions, and (3) satisfying conditions of applicability of certain statistical techniques. The application of transformations in each of these circumstances may be straightforward; nevertheless, account must always be taken of the degree of approximation desired or permissible. In addition, situations involving more than the single case of (1), (2) or (3) will frequently call for incompatible transformations. Such a case might arise, for example, in testing a theoretical prediction of linearity of relationship between two transformed variables; here it might be found that non-normality and heterogeneity

of variability around the theoretical line are introduced by the transforming operations. Such a state of affairs may make certain statistical tests inapplicable. Approximate methods are mentioned by which some problems of this sort may be solved. When sufficient information is available, these methods may be of value in determining the magnitudes of desired constants. Little is gained by employing these statistical methods of curve fitting if adequate information is not present.

For the theoretical psychologist many important considerations, other than those concerned with statistical tests, attach to the various kinds of distributions of two variables. Certain forms in which data may be expressed may lead us to as many "laws" as there are measures of "central value" of groups of data. In such cases appropriate transformations serve to reduce the multiplicity of forms of functional relation between variables, and these transformations provide both a subject matter and a tool for the theorist.

BIBLIOGRAPHY

1. BAKER, G. A. Transformation of non-normal frequency-distributions into normal distributions. *Ann. math. Statist.*, 1934, **5**, 113–123.
2. BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39–52.
3. BARTLETT, M. S. Square root transformations in the analysis of variance. *J. roy. statist. Soc., Suppl.*, 1936, **3**, 68.
4. BEALL, G. The transformation of data from entomological field experiments so that analysis of variance becomes applicable. *Biometrika*, 1942, **32**, 243–262.
5. CHAPMAN, R. A. Applicability of the z-test to a Poisson distribution. *Biometrika*, 1938, **30**, 188–190.
6. COCHRAN, W. G. The analysis of variance when the experimental errors follow the Poisson and binomial laws. *Ann. math. Statist.*, 1940, **11**, 335–347.
7. COCHRAN, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, **3**, 22–38.
8. CRAIG, A. T. On the distribution of certain statistics. *Amer. J. Math.*, 1932, **54**, 353–366.
9. CRAMER, H. *Mathematical methods of statistics*. Princeton, N. J.: Princeton Univ. Press, 1946.
10. CROZIER, W. J., & HOLWAY, A. H. On the law for minimal discrimination of intensities. III. *Proc. nat. Acad. Sci.*, 1938, **24**, 130–135.
11. CURTISS, J. H. On transformations used in the analysis of variance. *Ann. math. Statist.*, 1943, **14**, 107–122.
12. DIXON, W. J. A criterion for testing the hypothesis that two samples are from the same population. *Ann. math. Statist.*, 1940, **11**, 199–204.
13. EISENHART, C., HASTAY, M. W., & WALLIS, W. A. *Techniques of statistical analysis*. New York: McGraw-Hill, 1947.
14. EZEKIEL, M. *Methods of correlation analysis*. New York: John Wiley, 1930.
15. FELSINGER, J. M., GLADSTONE, A. I., YAMAGUCHI, H. G., & HULL, C. L. Reaction latency ($_st_r$) as a function of the number of reinforcements

(*N*). *J. exp. Psychol.*, 1947, **37**, 214–228.

16. Festinger, L. A statistical test for means of samples from skew populations. *Psychometrika*, 1943, **8**, 205–210.

17. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 97–105.

18. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.

19. Fry, T. C. *Probability and its engineering uses.* New York: Van Nostrand, 1928.

20. Graham, C. H., Brown, R. H., & Mote, F. A. The relation of size of stimulus and intensity in the human eye: I. Intensity thresholds for white light. *J. exp. Psychol.*, 1939, **24**, 555–573.

21. Graham, C. H., & Gagné, R. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, **26**, 251–280.

22. Haggard, E. A. Experimental studies in affective processes: II. On the quantification and evaluation of "measured" changes in skin resistance. *J. exp. Psychol.*, 1945, **35**, 46–56.

23. Haggard, E. A., & Garner, W. R. An empirical test of a derived measure of changes in skin resistance. *J. exp. Psychol.*, 1946, **36**, 59–70.

24. Hammer, Elaine R. Temporal factors in figural after-effects. *Amer. J. Psychol.* (In press.)

25. Hecht, S. The nature of the photoreceptor process. In C. Murchison (Ed.), *Handbook of general experimental psychology.* Worcester: Clark Univ. Press, 1934.

26. Hecht, S. A theory of visual intensity discrimination. *J. gen. Physiol.*, 1935, **18**, 767–789.

27. Holzinger, K. J., & Church, A. E. R. On the means of samples from a U-shaped population. *Biometrika*, 1928, **20**, 361–388.

28. Hotelling, H., & Pabst, M. R. Rank correlation and tests of significance involving no assumption of normality. *Ann. math. Statist.*, 1936, **7**, 29–43.

29. Jenkins, T. N. Facilitation and inhibition. *Arch. Psychol., N. Y.*, 1926, No. 86.

30. Kendall, M. G. *The advanced theory of statistics.* Vol. II. London: Charles Griffin, 1946.

31. Lipka, J. *Graphical and mechanical computation.* New York: John Wiley, 1918.

32. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

33. Morgan, C. T. The statistical treatment of hoarding data. *J. comp. Psychol.*, 1945, **38**, 247–256.

34. Mueller, C. G., & Richmond, P. N. (Unpublished.)

35. Paulson, E. On certain likelihood ratio tests associated with the exponential distribution. *Ann. math. Statist.*, 1941, **12**, 301–306.

36. Pitman, E. J. G. Significance tests which may be applied to samples from any population. *J. roy. statist. Soc., Suppl.*, 1937, **4**, 119–130.

37. Pitman, E. J. G. II. The correlation coefficient test. *J. roy. statist. Soc. Suppl.*, 1937, **4**, 225–232.

38. Pitman, E. J. G. III. The analysis of variance test. *Biometrika*, 1938, **29**, 322–335.

39. Rider, P. R. On small samples from certain non-normal universes. *Ann. math. Statist.*, 1931, **2**, 48–65.

40. Rider, P. R. On the distribution of the ratio of mean to standard devia-

tion in small samples from non-normal universes. *Biometrika*, 1929, **21**, 124–143.

41. Schoenfeld, W. N., Antoinitis, J. J., & Bersh, P. J. Unconditioned response rate of the white rat in a bar-pressing apparatus. *J. comp. physiol. Psychol.* (In press.)
42. Smith, J. G., & Duncan, A. J. *Sampling statistics and applications.* New York: McGraw-Hill, 1945.
43. Smith, J. R. Spatial and binocular effects in human intensity discrimination. *J. gen. Psychol.*, 1936, **14**, 318–345.
44. Stephan, F. E. Alternative statements of percentage data in the fitting of logarithmic curves. *J. Amer. statist. Ass.*, 1931, **26**, 58–61.
45. Wald, A., & Wolfowitz, J. On a test whether two samples are from the same population. *Ann. math. Statist.*, 1940, **11**, 147–162.
46. Wald, G. Area and visual threshold. *J. Gen. Physiol.*, 1938, **21**, 269–287.
47. Williams, C. B. The use of logarithms in the interpretation of certain entomological problems. *Ann. appl. Biol.*, 1937, **24**, 404–414.
48. Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.* (In press.)
49. Zubin, J. Note on a transformation function for proportion and percentages. *J. appl. Psychol.*, 1935, **19**, 213–220.

Received February 23, 1949.

HISTORICAL BEGINNINGS OF CHILD PSYCHOLOGY

WAYNE DENNIS
University of Pittsburgh

Scientific child psychology is commonly said to have begun with the publication of Preyer's *Die Seele des Kindes* (36) in 1882[1] and Hall's *The Contents of Children's Minds* (**20**) in 1883. Both of these were important contributions. Preyer's general account of mental development in the first four years was based upon careful and detailed observations of his son, to which were added data contributed by others and some comparative facts concerning behavioral development in animals. Hall's publication represented a very extensive attempt to determine the child's familiarity with a large number of relatively ordinary objects and concepts upon entrance into the first grade. This study of children's concepts was oriented, on the one hand, toward a description of the child's thought tendencies, and, on the other hand, toward the educational implications of the limited experience and understanding of the child upon his entrance into school.

Without doubt these two works are important landmarks in the development of child psychology. However, the history of science warns against attributing absolute originality or uniqueness to any contribution. Every development must be the result of a developmental process. The trends which led Preyer and Hall to observe and record child behavior attracted other men as well, and Preyer and Hall were not the first to carry scientific enterprise into the realm of human infancy.

Both of these men were well aware of the contributions of their predecessors. Preyer's book contains many references to the records of earlier observers, although, to be sure, none had left records as complete as was his own. Hall's study was undertaken as a repetition of an earlier study conducted in Berlin by Bartholomäi (**3**). Hall, like Preyer, did a much better job than his predecessors. The works of Preyer and Hall were outstanding not because they were novel in conception, but because they were executed with great ability and thoroughness.

It is the aim of this paper to survey the observational studies of normal child psychology up to the time of the publication of *Die Seele des Kindes* and *The Contents of Children's Minds.* It is believed that a survey of this sort will indicate the proper prominence of the two studies which are usually referred to as the first major studies in child psychology but will, at the same time, show that the early development of child psychology was continuous and gradual.

---

[1] The date of publication of this book is sometimes given as 1881. This seems to be due to the fact that Preyer, in prefaces to editions beyond the first, gave 1881 as the date of the first edition, and later writers have accepted the author's statement. Preyer's *preface* to the first edition is dated October 6, 1881, but a copy of the first edition in the Library of the Surgeon General's Office, Washington, D. C., bears a publication date of 1882 and bibliographical sources of that period give 1882 as the date of the book. It seems likely, therefore, that Preyer, in writing later prefaces, erroneously referred to the date of the first preface as the date of publication of the first edition.

Before beginning this survey, it will be advisable to attempt to define its scope. It undertakes to include only observational records of normal child behavior; it excludes theoretical treatments of the child. Since nearly every theory of human nature involves some consideration of childhood, to include general and theoretical material on the nature of the child would materially change the task which we have set for ourselves. A justification for excluding this material is found in the fact that theories of child nature have been given adequate historical treatment in several histories of philosophy and in histories of education, while no adequate survey of the early observational studies of children is available.

We have excluded also all publications relating to the behavior of abnormal children. The early medical literature on many types of abnormal children such as hydrocephalics, microcephalics, anencephalics, idiots, and cretins is very extensive. Many of the reports contain some general observations about behavior. The orientation of these reports, however, is on the whole medical rather than psychological, and it seems best to limit the present report to psychological studies of normal children.

In organizing an historical survey, a purely chronological arrangement of data may be followed, or a primary division into topics may be made, the historical sequence being followed within each division. In the present instance, a purely chronological treatment would result in the juxtaposition of entirely unrelated studies. It has, therefore, seemed best to subdivide the studies by subject matter before beginning the historical treatment.

The arrangement of the topics is more or less arbitrary. Since biographical records are among the earliest, and are also the most numerous it seems appropriate to place the infant biographies at the beginning of the survey.

BIOGRAPHICAL STUDIES

Tiedemann's account (47) of the behavioral development of his son was the first infant biography to be published. It appeared in 1787. The years covered by Tiedemann's scanty observations were 1781–1784.

Three years of occasional observations were reported within the space of 25 printed pages.

While Tiedemann was the first to publish a record of the development of an individual child, at least two other persons kept such a record at an earlier date than did Tiedemann. The two earlier diarists were Heroard and Pestalozzi. Heroard was a physician in the French court who was assigned to care for the health of the young prince who became Louis XIII. The physician began his duties at the birth of the prince in 1601 and continued them for many years. Heroard kept a journal in which he recorded the chief events in the court, his own activities, and also, in some detail the health and development of his young charge. Thus, while the developmental data concerning young Louis XIII are mixed with other matters, they are relatively complete. Heroard's journal, edited by Soulie and Barthelemy (**21**) was not published until 1868. It has furnished the basis of a recent book by Crump (**7**), entitled *Nursery Life 300 Years Ago*. Although Miss Crump cites some of the behavioral data, her chief interest lies in picturing the life of the court, rather than the development of the dauphin.

Pestalozzi, in 1774, kept for a period a diary in which he recorded his attempts at the education of his four-year-old son and made some record of his son's behavior. None of this diary seems to have been published in Pestalozzi's lifetime (1746–1827). Parts of it were contained in Niederer's *Notes on Pestalozzi* (**32**) whose date of publication was 1828.

Several studies of individual children made in the first half of the nineteenth century lay unpublished for many years after they were completed. Bronson Alcott kept for a while a record of the development of a daughter born in 1831. This record, still unpublished, is probably contained in notebooks left by Alcott (**30**). A brief summary of this record was presented by Talbot (**46**) in 1882.

Charles Darwin's diary of the development of his infant son, begun in 1840, was not printed until 1877 (**10**). Strumpell's longitudinal study of his child (**43**) was kept during the years 1846 and 1847, but did not appear until 1880.

The earliest baby biography to be published in English was by an American, Mrs. Emma Willard. It appeared as an appendix to an American edition of Madame Necker de Saussure's *Progressive Education*. Its date of publication is 1835 (**50**).

Sigismund's book (**41**), which contains some biographical material, appeared in 1856. Four other individual studies antedated Preyer's *Die Seele des Kindes* by one or two years. These reports were by Taine (**45**), Wyma (**51**), Sully (**44**), and Champneys (**5**).

The data just summarized indicate that at least twelve biographical records were kept before Preyer undertook his observations of his son. They indicate also an increase in frequency of records of this type in the years just preceding Preyer's study.

LANGUAGE STUDIES

In a sense, the reports to be listed under this heading might also be called biographical investigations, since all of the early researches on infant speech are longitudinal records of one or two subjects. Many of the baby biographies previously cited contain some account of language development. However, none of those mentioned in the preceding section was solely concerned with speech, whereas that topic is almost the exclusive interest of those to be discussed here. Schleicher (**39**), who wrote in 1861, seems to have been the first to record the speech development of a specific child. His report was followed fairly closely by those of Holden (**24**), Pollock (**33**), Vierordt (**49**), Egger (**12**), Schultze (**40**), and Humphreys (**25**). Since each of these was primarily a brief presentation of individual developmental data, further description seems unnecessary here.

Hun (**26**) described in some detail the invention of new words by a 4½-year-old girl who understood English quite well. This new language she taught to her 18-month-old brother.

NORMATIVE STUDIES

Several early investigations may be grouped together because they share the common characteristic of being interested in the typical performance of some age level, or of being interested in determining the average age of children at the onset of some developmental item. The earliest of these normative studies was that of Feldman (**13**), published in 1833. Feldman was a candidate for the M. D. degree. At that time, it was customary to require each candidate to prepare a short dissertation in Latin. Feldman chose to write concerning the normal functioning of the human body, and included under the heading of the normal state of the body in infants, data on the onset of walking and talking. He presumably obtained his data from mothers' reports, although his statement is not clear on this point. He states that he "observed" these events in the case of 35 infants, but it seems scarcely possible that that he could have had the facilities or the time to observe personally the beginning of speech and of upright locomotion in so many subjects. Feldman[2] writes:

---

[2] I am indebted to my wife for the translation of this work. The original monograph is available in the Library of the Surgeon General's Office, Washington, D. C.

> If you inquire at what time children attain the faculty of walking, that thing I have observed in the case of 35 infants. Of these, 6 were able to walk at the 11th and 12th months; however, the rest required 13 months, and some even 18 months or a small part of the 18th month in order that this feat might be accomplished without aid. All concerning whom I have spoken enjoyed the best of health. . . . I have found among the 35 infants that one of them attained the faculty of speaking the first word in the 14th month, 8 in the 15th month, 19 in the 16th month, 3 in the 17th month, one in the 18th month, and one in the 19th month. Therefore, the normal time for first speech in infants would seem to be the sixteenth month.

It will be seen that Feldman's presentation of his data in regard to walking is very incomplete, but it is clear that the majority of his subjects were one year of age or older before walking alone. His distribution of cases for the onset of speaking is more detailed, but he presents data on only 33 of his 35 cases. Nevertheless his study is notable for preceding by many years the next attempt to set up empirical standards for the onset of locomotion and speech.

The first person to apply a uniform behavioral test throughout a large part of the life span was Quetelet (**37**) whose work appeared in 1835. Quetelet is known for his contributions to methodology in the field of census records, social statistics and vital statistics. However, his interest extended beyond what is usually thought of as comprising these fields, and he presents in his *Physique Sociale* observations on strength of upright pull and strength of grip of the right and the left hands from ages 6 to 60. His data are presented separately for the two sexes. Yearly age groups are employed from 6 to 21 years; there is a 25-year-old group; thereafter his subjects are grouped by decades. In each group, his averages are based upon ten persons of each sex. His data show that males are superior in strength to females at every age level. The data also show the superior strength of the right hand as compared to the left in dextral individuals.

In addition, Quetelet presented data on the incidence of suicide and of crimes at various epochs of life, including the period below ten years. The data are presented by 5-year step-intervals from 10 to 40 years, and by decades thereafter. Especially noteworthy is the high suicide rate of Berlin youth between 15 and 20 years of age. The data for Berlin refer to the years 1818–1824.

The first normative study of the newborn was that of Kussmaul (**28**), which appeared in 1859. Kussmaul, like Feldman, was a physician. His experiments with various modes of stimulating the newborn led him to believe that all of the sense organs were capable of some degree of function at birth. He gave fairly adequate description of the repertory of responses of the neonate. This pioneer study was followed by two similar investigations by Genzmer (19) and Kroner (27), which added but little to the findings of Kussmaul.

Other studies of neonatal behavior may be called normative only by extending somewhat the meaning of norms, yet it is true that these studies were designed to investigate the characteristics of an age period. Hertz (22) and Biedert (4) were interested in measuring the force of suction, or negative pressure, of the neonate's nursing response and devised apparatus which they used for that purpose. Donders (11), Raehlmann and Witkowski (38) and Cuignet (8) made observations upon the ocular responses of the neonate. Moldenhauer (31) tested the hearing of the young infant.

THE CONTENTS OF CHILDREN'S MINDS

The studies to be reviewed here served as models for Hall's work by the same title, and Hall in his introduction to his investigation gives an excellent account of his predecessors' work.

The first study, which was conducted in Berlin, was reported by Bartholomäi (3). While the Berlin investigators described their problem roughly as that of determining the contents of children's minds, the tests or questions which they submitted to their subjects called for several types of performance. One procedure consisted in asking each subject which of a number of common objects, such as a pond, a lake, a running hare, a squirrel in a tree, a hen with chickens, etc., the child had ever seen. In this portion of the study it was undetermined whether the child who had not had these experiences might nevertheless have some comprehension of the phenomena gained from pictures or from verbal descriptions; the child was interrogated only with regard to his direct experience with the objects. On the other hand, some questions which were employed were essentially tests of language comprehension. It was determined, for instance, whether or not the child knew what was meant by a dwelling, a sphere, a cube, clouds, etc. Other questions asked the child to give the name and business of his father. The examining was done by the regular classroom teachers to whom a circular or directions had been sent. Clearly the research procedures were far from ideal.

In this investigation, which was conducted in 1869 and whose results were published in 1870, 2238 children just entering school were questioned in regard to 75 topics. The chief discovery was the remarkable extent to which first graders are ignorant of simple phenomena. Less than 50% knew what was meant by a sunset, only 30% said they had seen a sunrise, only 11% had seen a river, etc. It was found that girls excelled in regard to knowledge of some concepts, boys in regard to others. Children who had attended kindergarten, on the whole, did better than those who had not attended, a finding which foreshadows an interest of the present day.

Lange (**29**), in 1879, presented similar questions to 500 children entering the city schools of Plauen, Germany, and to 300 children in near-by rural districts. In regard to the concepts which he attempted to test, many of which were of the nature study sort, the country children proved to have a superior knowledge.

G. Stanley Hall's study of Boston children (**20**) was begun in September, 1880. It was first published in the *Princeton Review* for 1883, and was later republished several times. Hall drew up a list of 134 topics about which his subjects were to be interrogated. It is not within the province of this article to describe Hall's study in full. We may note in passing that Hall's contribution in this field was chiefly that of improvement in method. Instead of utilizing ordinary classroom teachers whose methods of examination could not be controlled, Hall employed four kindergarten teachers who were trained to uniform procedures of questioning, and who frequently met with their director for discussion. In Boston, as in Berlin, the percentage of ignorance concerning simple concepts was surprisingly high.

GALTON'S "INQUIRIES INTO THE HUMAN FACULTY AND ITS DEVELOPMENT"

That Galton's work is not usually given a prominent place in the development of child psychology is probably due to the fact that this field was not his primary interest. It is true, nevertheless, that he should be credited with several contributions to this area of specialization. Although several of his writings bearing upon childhood were included in the *Inquiries Into the Human Faculty and Its Development* which was published in 1882, their first date of publication in most instances was earlier than that year, since Galton's book consisted for the most part of reprints of articles that had previously been presented in scientific and popular journals.

In 1875 Galton (**14**) published the first scientific study of the psychological development of twins, thus initiating one of the major types of investigations of nature and nurture. Galton's method consisted in inquiring concerning twins who in childhood were very much alike to determine whether or not they became more unlike after they left their homes and entered different environments. Conversely, he sought information concerning twins who in the earliest years were exceedingly different, to learn whether or not under the continued influence of the same nurture they became more similar. His materials were obtained by sending "circulars of inquiry" to persons who were either twins themselves or were relatives of twins.

Galton found in regard to these twins who were remarkably alike in childhood that in some cases "the resemblance of body and mind continued unaltered up to old age, notwithstanding very different conditions of life." Other cases showed divergence with age. He felt that the continued close similarity of some cases despite different environments showed that the divergences of other twins were due to illnesses or to different original endowments which only became apparent in maturity. The second group of data, derived from twins who were originally dissimilar, likewise led Galton to stress the importance of nature, since he found that no matter how similar the environment of the two, they remained markedly different. Galton wrote: "There is no escape from the conclusion that nature prevails enormously over nurture when the differences of nurture do not exceed what is commonly to be found among persons of the same rank of society and in the same country."

In Galton's "Psychometric experiments," first presented in 1879, (**15**), he employed himself as subject in studying the association of ideas. It may seem odd, therefore, to cite this research as a contribution to child psychology. One of Galton's tasks, however, was to attempt to determine when each association had been formed. In many cases this was not difficult, since the date of formation of the association could be established with certainty. Galton found that of 124 associations, 39% of them were formed in boyhood and youth, 46% in adult years, and only 15% could be attributed to recent events. On the basis of these results Galton called attention to the importance of childhood in the formation of associations between ideas.

"Statistics of mental imagery" was the title of a report by Galton in 1880 (**16**). In pursuing his interest in imagery, Galton, as is well-known, devised rating scales for several aspects of imagery. Visual imagery was rated by 100 English men of science and by 172 Charterhouse school boys. Vivid imagery was found to be much more common among the boys, and boys were more able also to call up colored images. Galton believed that the habit of abstract thought resulted in the suppression of visual images.

The study of imagery was pursued further by Galton in an investigation of number forms (**17**). He found that not uncommonly a person located each number within a certain region of his visual space. Most of his informants were certain that their particular forms of visualizing numbers had been present, unchanged, from early childhood. This finding led Galton to ask teachers to question their pupils concerning this matter. Among subjects of about 14 and 15 years of age, Galton found that clear-cut number forms appeared in about one case in 20 in boys and more often in girls.

On the basis of the number of contributions to the study of mental development which are contained in *Inquiries Into the Human Faculty and Its Development*, it is our belief that this publication should be grouped with those of Preyer and Hall as one of the important early publications in child psychology.

MISCELLANEOUS STUDIES

In addition to the studies previously discussed, which were capable of grouping, a few remain, each of which stands alone within the epoch under consideration. We shall mention these briefly, following the chronological order.

Miss Elizabeth Peabody (1) in 1835, 1836 and 1837 published accounts of the procedures of Bronson Alcott in his remarkable Boston school. Alcott stimulated student participtiion in the educational situation by holding "conversations" with children. Miss Peabody recorded, as nearly verbatim as possible, many of these conversations. The result was a detailed account of child reasoning and child response in Alcott's school.

Darwin's "Expression of the Emotions in Animals and Man" (9) contained some observational material on emotional expression in children. Many of these observations were made on his own children; we have seen that his interest in child behavior began as early as 1840. Some observations were gathered from friends and correspondents.

Clarke (6) described certain images which commonly appear in childhood but which in most cases disappear with age. It seems likely that he was referring to what are now known as eidetic images.

Sikorsky (42) was the first to study decrements of performance during the school day. Investigations of this type later came to be known as studies of mental fatigue.

Vierordt (48) devised a shoe which would record the contact of the feet with the walking surface, and pointed out differences between infantile and adult walking patterns.

Hicks (23) devised a method of recording prenatal movements by measuring displacements of the abdominal wall.

PROFESSIONAL AFFILIATIONS OF EARLY RESEARCH WORKERS IN THIS FIELD

It may be of interest to note the professional affiliations of those who first contributed to our knowledge of child behavior. The occupations of some of the persons to whom we have referred could not be ascertained from the sources available to us. However, it is obvious that physicians were the most frequent contributors to the as-yet-unselfconscious science. Among the members of the medical profession mentioned in our bibliography are Biedert, Clark, Cuignet, Feldman, Genzmer, Héroard, Hicks, Hun, Kroner, Kussmaul, Moldenhauer, and Wyma. Educators are next in frequency. This professional group includes Alcott and Peabody, Bartholomäi, Lange, Pestalozzi and Strumpell. Vierordt was a physiologist; Darwin a biologist; Quetelet was many things, but was primarily a statistician; Taine and Sully both were philosopher-psychologists. Galton contributed to so many fields that he can scarcely be claimed by any. Preyer was a physiologist. Hall belongs to both psychology and education.

DISCUSSION AND SUMMARY

We have reviewed forty-two publications which appeared before 1882. It will be noted that the earliest publication referred to in the preceding survey is the biographical record of Tiedemann, which appeared in 1787. Tiedemann deserves to be credited with a very real scientific priority. His record of his son's behavioral development was thoroughly scientific in spirit. Although Héroard's record precedes Tiedemann's, Héroard's interest was that of a diarist and historian; his account of the behavior of young Louis XIII was not oriented toward the study of human development.

Following Tiedemann, there seems to be a gap of forty-one years during which we have no record of any publication relating to normal child behavior. In 1828, Niederer published part of Pestalozzi's notes of a father. Thereafter, there was a consistent increase in publication. Between 1830 and 1860 there were seven publications. The 1860's saw the appearance of four titles. A sharp increase in publication occurred during the 1870's, the total number during that decade being seventeen. During the years 1880 and 1881 alone, 12 titles are listed in our bibliography.

The decade by decade comparison of publication indicates that an increase in activity toward the development of modern child psychology had been under way for some time before the publications of Preyer and Hall. These two scientists were apparently responding to the same influences of the times as were lesser men, but since Preyer and Hall were abler workers their contributions were more outstanding.

In this survey we have seen no cause to question the view that Preyer's *The Mind of the Child* and Hall's *Contents of Children's Minds* were the most notable contributions appearing before 1885. It has been suggested, however, that Galton's *Inquiry into the Human Faculty and Its Development* should be rated as almost equal to the afore-mentioned publications of Preyer and Hall in regard to its contributions to child psychology.

this matter. Among subjects of about 14 and 15 years of age, Galton found that clear-cut number forms appeared in about one case in 20 in boys and more often in girls.

On the basis of the number of contributions to the study of mental development which are contained in *Inquiries Into the Human Faculty and Its Development*, it is our belief that this publication should be grouped with those of Preyer and Hall as one of the important early publications in child psychology.

MISCELLANEOUS STUDIES

In addition to the studies previously discussed, which were capable of grouping, a few remain, each of which stands alone within the epoch under consideration. We shall mention these briefly, following the chronological order.

Miss Elizabeth Peabody (1) in 1835, 1836 and 1837 published accounts of the procedures of Bronson Alcott in his remarkable Boston school. Alcott stimulated student particiption in the educational situation by holding "conversations" with children. Miss Peabody recorded, as nearly verbatim as possible, many of these conversations. The result was a detailed account of child reasoning and child response in Alcott's school.

Darwin's "Expression of the Emotions in Animals and Man" (9) contained some observational material on emotional expression in children. Many of these observations were made on his own children; we have seen that his interest in child behavior began as early as 1840. Some observations were gathered from friends and correspondents.

Clarke (6) described certain images which commonly appear in childhood but which in most cases disappear with age. It seems likely that he was referring to what are now known as eidetic images.

Sikorsky (42) was the first to study decrements of performance during the school day. Investigations of this type later came to be known as studies of mental fatigue.

Vierordt (48) devised a shoe which would record the contact of the feet with the walking surface, and pointed out differences between infantile and adult walking patterns.

Hicks (23) devised a method of recording prenatal movements by measuring displacements of the abdominal wall.

PROFESSIONAL AFFILIATIONS OF EARLY RESEARCH WORKERS IN THIS FIELD

It may be of interest to note the professional affiliations of those who first contributed to our knowledge of child behavior. The occupations of some of the persons to whom we have referred could not be ascertained from the sources available to us. However, it is obvious that physicians were the most frequent contributors to the as-yet-unselfconscious science. Among the members of the medical profession mentioned in our bibliography are Biedert, Clark, Cuignet, Feldman, Genzmer, Héroard, Hicks, Hun, Kroner, Kussmaul, Moldenhauer, and Wyma. Educators are next in frequency. This professional group includes Alcott and Peabody, Bartholomäi, Lange, Pestalozzi and Strumpell. Vierordt was a physiologist; Darwin a biologist; Quetelet was many things, but was primarily a statistician; Taine and Sully both were philosopher-psychologists. Galton contributed to so many fields that he can scarcely be claimed by any. Preyer was a physiologist. Hall belongs to both psychology and education.

DISCUSSION AND SUMMARY

We have reviewed forty-two publications which appeared before 1882. It will be noted that the earliest publication referred to in the preceding survey is the biographical record of Tiedemann, which appeared in 1787. Tiedemann deserves to be credited with a very real scientific priority. His record of his son's behavioral development was thoroughly scientific in spirit. Although Héroard's record precedes Tiedemann's, Héroard's interest was that of a diarist and historian; his account of the behavior of young Louis XIII was not oriented toward the study of human development.

Following Tiedemann, there seems to be a gap of forty-one years during which we have no record of any publication relating to normal child behavior. In 1828, Niederer published part of Pestalozzi's notes of a father. Thereafter, there was a consistent increase in publication. Between 1830 and 1860 there were seven publications. The 1860's saw the appearance of four titles. A sharp increase in publication occurred during the 1870's, the total number during that decade being seventeen. During the years 1880 and 1881 alone, 12 titles are listed in our bibliography.

The decade by decade comparison of publication indicates that an increase in activity toward the development of modern child psychology had been under way for some time before the publications of Preyer and Hall. These two scientists were apparently responding to the same influences of the times as were lesser men, but since Preyer and Hall were abler workers their contributions were more outstanding.

In this survey we have seen no cause to question the view that Preyer's *The Mind of the Child* and Hall's *Contents of Children's Minds* were the most notable contributions appearing before 1885. It has been suggested, however, that Galton's *Inquiry into the Human Faculty and Its Development* should be rated as almost equal to the afore-mentioned publications of Preyer and Hall in regard to its contributions to child psychology.

BIBLIOGRAPHY

1. Alcott, B., & Peabody, Elizabeth P. *Record of a school.* Boston, N. Y., Phila.: James Munroe & Co., 1835.
2. Alcott, B., & Peabody, Elizabeth P. *Conversations with children on the Gospels.* Boston: James Munroe & Co., Vol. 1, 1836, Vol. 2, 1837.
3. Bartholomäi, F. The contents of children's minds on entering school at the age of 6 years. *Städtisches Jahrbuch, Berlin und Seine Entwicklung,* 1870, **4.** (Translation published in Rep. U. S. Com. Educ. for year 1900–1901, **1,** 709–729. Washington: U. S. Government Printing Office, 1902.)
4. Biedert, P. Zur Technik der Pneumatometrie. *Dtsch. arch. f. klin. med.* 1875, **18,** 115–117.
5. Champneys, F. H. Notes on an infant. *Mind,* 1881, **6,** 104–107.
6. Clarke, E. H. *Visions: a study of false sight.* Boston: Houghton, Osgood, & Co., 1878.
7. Crump, L. *Nursery life 300 years ago.* London: Geo. Routledge & Sons. 1929.
8. Cuignet, L. De la vision chez le tout jeune enfant. *Ann. Ocul.,* 1871, **66,** 117–126.
9. Darwin, C. *The expression of the emotions in man and animals.* New York: Appleton, 1873.
10. Darwin, C. Biographical sketch of an infant. *Mind,* 1877, **2,** 285–294.
11. Donders, F. C. Versuch einer genetischen Erklärung der Augen-Bewegungen. *Pflüg. Arch. f. d. ges. Physiol.,* 1876, **13,** 373–421.
12. Egger, M. E. *Observations et reflexions sur le developpement de l'intelligence et du langage chez les enfants.* Paris: Picard, 1870.
13. Feldman, Henricus. *De statu normali functionum corporis humani animadversiones quandam.* Bonne: C. Georgie, 1833.
14. Galton, F. Twins, as a criterion of the relative power of nature and nurture. *J. anthrop. Inst.,* 1875, **5,** 324–329.
15. Galton, F. Psychometric experiments. *Brain,* 1879, **2,** 149–157.
16. Galton, F. Statistics of mental imagery. *Mind,* 1880, **5,** 301–318.
17. Galton, F. Visualized numerals. *J. Roy. anthrop. Inst.,* 1880, **10,** 85–102.
18. Galton, F. *Inquiries into the human faculty and its development.* London: Macmillan, 1883.
19. Genzmer, A. *Untersuchungen über die Sinneswahrnehmungen des neugeborenen Menschen.* Halle: Max Niemeyer, 1882.
20. Hall, G. S. Contents of children's minds. *Princeton Rev.,* 1883, **11,** 249–272.
21. Héroard, Jean. Journal, 1601–1628. Ed. by E. Soulie and E. de Barthelemy, 1868.
22. Hertz, M. Einige Bermerkungen über das Saugen der Kinder. *Jahrb. f. Kinderhk.,* 1865, **7,** 46–57.
23. Hicks, J. B. On recording the fetal movements by means of a gastrograph. *Trans. obstet.* Soc. *London,* 1880, **22,** 134.
24. Holden, E. S. On the vocabularies of children under two years of age. *Trans. Am. philol. Assoc.,* 1877, 58–68.
25. Humphreys, M. W. A contribution to infantile linguistics. *Trans. Am. philol. Ass.,* 1880, **11,** 5–17. (Hartford: Case, Lockwood & Brainard Co., 1880.)
26. Hun, E. R. Singular development of language in a child. *Quart. J. psychol. Med.,* 1868, **2,** 525–528.
27. Kroner, T. Über die Sinnesempfindungen der Neugeborenen. *Breslauer aerzt. Zschl.,* 1882, **4,** 37–41.
28. Kussmaul, A. *Untersuchungen über das Seelenleben des neugeborenen Menschen.* Heidelberg & Leipzig: C. F. Winter, 1859.

29. Lange, K. Der Vorstellungskreis unserer sechsjahrigen Kleinen. *Allgemeine Schultzeitung*, 1879, **56**, 327ff.

30. McCuskey, D. *Bronson Alcott, teacher.* New York: Macmillan, 1940.

31. Moldenhauer, W. Zur Physiologie des hörogans neugeborener. *Beitrage zur Geburtshülfe, Gynäkologie und Pädriatrik, Leipzig.* 1881, 199–204.

32. Niederer, Johannes. *Notes on Pestalozzi.* Aix-la-Chapelle, 1828. (Cited by R. De Guimps, *Pestalozzi, his life and work.* New York: Appleton, 1906. Pp. 40–48.)

33. Pollock, F. An infant's progress in language. *Mind*, 1878, **3**, 392–401.

34. Pollock, F. Review of Schulte's 'Die Sprache des Kindes.' *Mind*, 1881, **6**, 436–438.

35. Preyer, Wilhelm. Psychogenesis. *Deutsche Rundschau*, 1880, **23**, 198–221. (Translated by Marion Talbot, *J. spec. Philos.*, 1881, **15**, 159–188.)

36. Preyer, W. *Die Seele des Kindes. Beobachtungen über die geistige Entwicklung des Menschen in den ersten Lebensjahren.* Leipzig: T. Grieben, 1882.

37. Quetelet, L. A. J. *Sur l'homme et le developpement des ses facultés, ou essai de physique sociale.* Paris: Bachdier, 1835.

38. Raelhmann, E., & Witkowski, L. Über atypische Augenbewegungen. *Arch. f. Anat. u. Physiol.*, 1877, **1**, 454–471.

39. Schleicher, A. Einige Beobachtungen an Kindern. *Beitrage zur vergl. Sprachforsh.*, 1861, **2**, 497–498.

40. Schultze, F. *Die Sprache des Kindes. Eine Anregung zur Erforschung des Gegenstandes.* Leipzig: Günther, 1880.

41. Sigismund, Berthold. *Kind und Welt. Vatern, Muttern und Kindefrunden gewidmet.* Braunschweig: Vieweg, 1856.

42. Sikorsky, A. J. Sur les effets de la lassitude provoquée par les travaux intellectuels chez les enfants de l'âge scolaire. *Ann. d'hygiene publ. Paris*, 1879, **2**, series 3, 458–464.

43. Strumpell, L. Notizen über die geistige Entwicklung eines weiblichen Kindes Während der ersten zwei Lebenjahr. In *Psychologische Pedagogik*, Leipzig: Georg Bohme, 1880.

44. Sully, J. Babies and science. *Cornhill Mag.*, 1881, **43**, 539–554.

45. Taine, H. Note sur l'acquisition de langage chez les enfants et dans l'espèce humaine. *Rev. phil.*, 1876, **1**, 3–23. (Translated in *Mind*, 1877, **2**, 252–257.)

46. Talbot, E. *Papers on infant development.* Educ. Dept., Amer. Soc. Sci. Assoc., 1882, pp. 52.

47. Tiedemann, D. Beobachtungen über die Entwicklung der Seelenfahrigkeiten bei Kindern. *Hessichen Beitrage zur Gelehrsamkeit und Kunst*, 1787, **2**, 313 ff.; **3**, 486 ff. (Translated by S. Langer and C. Murchison, *J. genet. Psychol.*, 1927, **34**, 205–230.)

48. Vierordt, K. *Der Gang des Menschen.* Tubingen: Laupp, 1881.

49. Vierordt, K. Die Sprache des kindes. *Dtsche. Rev.*, 1879, **4**, 29–46.

50. Willard, Mrs. Emma. Observations upon an infant during its first year. By a mother. In *Progressive Education*, by Madame Necker de Saussure. Boston: Wm. D. Ticknor, 1835, pp. 323–348.

51. Wyma, ———. The mental development of the infant of today. *J. psychol. Med. & mental Pathol. New series*, 1881, **7**, 62–69.

Received March 3, 1949.

WIENER, NORBERT. *Cybernetics* (*or control and communication in the animal and the machine*). New York: John Wiley, 1948. Pp. ii+194.

This curious and provocative book is actually a series of essays on a variety of topics rather than an integrated work. It treats of such diversities as servo mechanisms, statistical mechanics, time, digital and analogue computers, physiology of the nervous system, perception, psychopathology, prosthesis, and the nature of society. On each of these seemingly unrelated subjects Dr. Wiener has something challenging to say. These concepts and subject matters represent to the author sub-topics in the new discipline of cybernetics. This term, which is formed from the Greek word meaning *steersman*, is defined as embracing the entire field of control and communication theory as applied to both animals and machines.

The author is professor of mathematics at the Massachusetts Institute of Technology. It is not surprising, therefore, to find that approximately one third of the book treats the sophisticated mathematics for which he is famous. In this part of the work, which will probably be fully understood only by professional mathematicians, the author develops the communication theory which he proposes as a basic theoretical framework for treating the activity of complex machines and living organisms. Beginning this portion with a brief history of the statistical mechanics of Gibbs and Lebesgue, Dr. Wiener discusses the development of ergodic theory under Koopman and von Neumann, then speaks briefly of entropy and the Maxwell demon and finally passes on to the development of a statistical mechanics of time series. It is the latter which the author regards as most appropriate in discussing animal behavior.

Fortunately for the majority of readers, mathematical notation is held to a minimum in the remainder of the book. Sometimes with great simplicity, but always brilliantly, the author hurries from discussions of the concept of time, through rather technical yet fascinating discussions of servo mechanisms and electronic computers, through essays on psychopathology and human perception to pessimistic prognostications concerning the fate of man and his society. One is a little breathless when he arrives at the end of the last chapter where, incidentally, he discovers a final note on how to construct a chess playing machine.

There are many challenging speculations in this work which will provoke discussions and, it is to be hoped, research. But the author's most important assertion for psychology is his suggestion, often repeated, that computers, servos, and other machines may profitably be used as models of human and animal behavior. Although it is nowhere stated explicitly, the implication is strong that the value to the physiologist and psychologist of these physical models lies less within themselves than in the ready-made mathematical theory which is now associated with them. It seems that Dr. Wiener is suggesting that psychologists should borrow the theory and mathematics worked out for machines and apply them to the behavior of men.

Action upon such a proposal is, of course, contingent upon evidence of the degree of similarity between the analogue and that to which it is supposed to be analogous. A close correspondence between the two would suggest that there might be some profit in applying the mathematics of the one to the other, though no degree of correspondence short of perfect identity could relieve the researcher of his scientific caution.

Obviously, since men can not as yet fabricate themselves out of vacuum tubes and circuits there is less than perfect identity between man and the models proposed by Dr. Wiener. But is the parallel sufficiently close to be encouraging?

Little evidence on this point is cited in the book other than examples of gross similarity between the action of men and machines. Indication is given of a parallel between voluntary activity and the response of a closed loop servo which responds, not to an absolute signal, but to the difference between a desired state of affairs and the state of affairs which exists at the moment. Certain forms of ataxia are recognized as suggesting certain conditions of servo instability, while others are said to resemble the response of a servo with an open feedback loop. The activity of the brain is compared to the working of an electronic computer and the suggestion is made that certain forms of human insanity have their parallel in computer breakdowns.

These and other like observations constitute almost the entire body of "evidence" that machines closely resemble men in their activities. Almost no factual data other than of a purely physiological nature exist to lend credence to this important postulate.

In fairness to the author it must be pointed out that no attempt is made to "prove" the fruitfulness of the proposed analogous reasoning nor to provide an exhaustive catalogue of man-machine correspondences. Dr. Wiener is content to develop the mathematical theory of machines and then to point out certain similarities between the responses of these devices and those of living beings. He may regard it as the task of others to fill in the details of the pattern which he has sketched in outline.

Yet one cannot help but regret that more of the details were not at hand when the book was written. In the final analysis it is precisely these particulars which will determine whether the book should be judged as literature or science. At present one can only hope that cybernetics will prove to be as fruitful as it now stimulating to the imagination.

FRANKLIN V. TAYLOR.

Naval Research Laboratory.

BOOKS AND MATERIALS RECEIVED

Andrews, T. G., *Some psychological apparatus: A classified bibliography.* Psychological Monographs, No. 289. Washington, D. C.: American Psychological Assn.: 1948. Pp. 38.

Ansbacher, H. L. *Attitudes of German prisoners of war: a study of the dynamics of national-socialistic followership.* Psychological Monographs, No. 288. Washington, D. C.: American Psychological Assn. 1948. Pp. 42. $1.00

Blum, Milton L. *Industrial psychology.* New York: Harper, 1949. Pp. xi+519. $4.50.

Bouthilet, Lorraine, and Byrne, Katherine Mann. *You and your mental abilities.* Chicago: Science Research Associates, 1948. Pp. 48.

Breckenridge, Marian E., and Vincent, E. Lee. *Child development.* Philadelphia: W. B. Saunders, 1949. Pp. 622. $4.00.

Brouwer, Paul J. *Student personnel services in general education.* Washington, D. C.: American Council on Education, 1949. Pp. xix+317. $3.50.

Buros, Oscar K. *The third mental measurements year book.* New Brunswick: Rutgers Univ. Press, 1949. Pp. xiv+1246. $12.50.

Das, Sonya Ruth. *The American woman in modern marriage.* New York: Philosophical Library, 1948. Pp. 185. $3.75.

De Ford, Miriam Allen. *Psychologist unretired.* Stanford Univ., Calif.: Stanford Univ. Press, 1948. Pp. ix+127. $3.00.

De Grazia, Sebastian. *The political community.* Chicago: Univ. of Chicago Press, 1948. Pp. xx+258. $4.00.

Erickson, Clifford E. *A practical handbook for school counselors.* New York: Ronald Press, 1949. Pp. vii+224. $3.00.

Escalona, Sibylle K. *An application of the level of aspiration experiment to the study of personality.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1948. Pp. viii+132. $2.10.

Haring, Douglas G. *Personal character and cultural milieu.* (Rev. Ed.) New York: Syracuse Univ. Press, 1949. Pp. x+670. $5.00.

Harris, Noel G. *Modern trends in psychological medicine.* London: Paul B. Hoeber, 1948. Pp. xii+450. $10.00.

Havighurst, Robert J., and Taba, Hilda. *Adolescent character and personality.* New York: John Wiley, 1949. Pp. x+315. $4.00.

Hollingworth, Harry L. *Psychology and ethics.* New York: Ronald Press, 1949. Pp. ix+247. $3.50.

Husband, Richard W. *Applied psychology.* New York: Harper, 1949. Pp. xiii+845.

Jennings, Helen Hall. *Sociometry in group relations.* Washington, D. C.: American Council on Education, 1948. Pp. ix+85. $1.25.

Kanner, Leo. *Child psychiatry.* (2nd Ed.) Springfield, Ill.: Charles C Thomas, 1949. Pp. xxiv+752. $8.50.

Klein, Melanie. *Contributions to psychoanalysis.* London: Hogarth Press, 1948. Pp. 416.

Kunst, Mary S. *A study of thumb- and finger-sucking in infants.* Psychological Monographs, No. 290. Washington. D. C.: American Psychological Assn., 1948.

Lennep, D. J. Van. *Four picture test.* The Hague: Martinus Nijhoff, 1948. Pp. 48.

McNemar, Quinn. *Psychological statistics.* New York: John Wiley, 1949. Pp. vii+364. $4.50.

Machover, Karen. *Personality projection.* Springfield, Ill.: Charles C Thomas, 1949. Pp. ix+191. $3.50.

Menninger, William C. *Understanding yourself.* Chicago: Science Research Associates, 1948. Pp. 52.

Miner, Roy Waldo, *et al. Newer synthetic analgesics.* New York: New York Academy of Sciences, 1948. Pp. 174.

Murphy, Gardner. *Historical introduction to modern psychology.* (Rev. Ed.) New York: Harcourt, Brace, 1949. Pp. xiv+466.

Rotman, David B. *Addresses and papers.* Chicago: Psychiatric Institute, 1948. Pp. xiv+204.

Sadler, William S. *Adolescence problems.* St. Louis: C. V. Mosby, 1948. Pp. 466. $4.75.

Schiller, Paul V. *Aufgabe der psychologie.* Wein: Springer Verlag, 1948. Pp. 233.

Segel, David. *Intellectual abilities in the adolescent period.* Washington: Federal Security Agency (Office of Education), Bulletin 1948, No. 6.

Smith, Gudmund. *Psychological studies in twin differences.* Lund, Sweden: C. W. K. Gleerup, 1949. Pp. 251.

Swineford, Frances. *A study in factor analysis.* Chicago: Univ. of Chicago Press, 1948. Pp. xi+70. $2.00.

Symonds, Percival M. *Symond's picture-story test.* New York: Bureau of Publications, Teachers College, Columbia Univ. 1948. Pp. 20. $5.25.

Symonds, Percival M. *Dynamic psychology.* New York: Appleton-Century-Crofts, 1949. Pp. vii+413. $3.75.

GUIDING HUMAN MISFITS

By ALEXANDRA ADLER, *M.D.*

Asst. Clinical Prof. of Neurology, N.Y.U. College of Medicine

THE author's ideas here set forth, in the light of her own clinical experience, follow closely those which were laid down by her father, Alfred Adler, the founder of individual psychology. She herself has specialized in neurology, psychiatry and psychotherapy, both in America and abroad. It should be helpful to physicians, social workers and teachers, as well as to educated parents and other lay people. $2.75

THE EMOTIONS:

Outline Of A Theory

By JEAN-PAUL SARTRE

Analyzes the roles which fear, lust, melancholy and anguish play in the life of man, and what is the true reality of conscious life. $2.75

THE PSYCHOLOGY OF IMAGINATION

By JEAN-PAUL SARTRE

Describes the great "unrealizing" function of the mind. By means of a close consideration of the operations of imagination and of the nature of the imaginary, Sartre reveals a new way of conceiving the nature of the psychic life and of the mind's relationship with the external world. $3.75

At Your Bookstore, or Use
Order Coupon

PHILOSOPHICAL
LIBRARY
Publishers

PHILOSOPHICAL LIBRARY, Publishers
15 East 40th St., Dept. 82
New York 16, N.Y.
Please send me .... copies of (write in margin) at $................. per copy. Enclosed are $.......................................
NAME ....
ADDRESS ....
(Expedite shipment by enclosing remittance.)

REPORT TO LIBRARIANS

Available Back Issues of the Psychological Bulletin

Thirty-three issues of the Psychological Bulletin became out-of-print during 1948. Libraries and institutions ordering "All Available Issues" ordered 10 sets of this journal.

Estimating a similar number of sales of all available issues during 1949, the inventory shows that 17 other issues will become out-of-print during the next year; that is, there are as of January 2, 1949, 17 issues of fewer than ten copies each.

Copies of this journal are stored in the warehouses of the two printers who have printed it between 1904 and 1949. About 390 separate issues are handled for an "All Available Issues" sale. Three weeks should be allowed as a minimum before expecting such an order to be filled. For quicker handling, the APA keeps issues of the last two years available in the APA office.

American Psychological Association
1515 Massachusetts Avenue Northwest
Washington 5, D.C.

Vol. 46, No. 4 July, 1949

Psychological Bulletin

A HISTORICAL SURVEY OF PSYCHOLOGY IN THE SOVIET UNION[1]

IVAN D. LONDON
Tulane University

To trace the historical development of psychology in the Soviet Union is to acquire the means for an understanding of its present status and content, about which so little is known outside the borders of the Soviet Union. In this paper it is proposed to survey the historical antecedents of contemporary dialectic psychology in the Soviet Union and to provide some idea of its basic tenets as presently constituted. While it were best to give a historical exposition which could take us down to the present, the purposes of this paper, as it so happens, do not necessitate a detailed account of events in the Soviet psychological world beyond 1936. The writer hopes, however, to supplement the present paper in the future with a historical account of developments subsequent to 1936 and so help overcome the unfortunate meagerness in knowledge of the workings and level of accomplishment of dialectic psychology in the Soviet Union.

When one surveys from the vantage point of the present the course that psychology in the Soviet Union has taken since the days of the October Revolution, it is possible to mark out three periods in its evolution. These periods in the evolution which eventuated in the establishment of a dialectical psychology, distinct from all other schools of psychology, may be designated roughly as (1) the pre-Marxian period (1917 to 1923 inclusive), (2) the period of trial and "error"[2]

---

[1] The writing of this paper was initiated in the course of independent study on Soviet psychology at Northwestern University. All direct quotations from Russian sources are the writer's translations. The writer wishes also to bring to the reader's attention the fact that, wherever a Soviet psychologist's point of view or argument is presented, an attempt has been made to reproduce or, at least, to convey the usual tone and style of such presentation as far as it is possible to do so in English.

[2] The quotes here and elsewhere harbor no invidious or any other intent. Context should in every instance explain their use.

(1924 to 1930–1931), and (3) the period of the gradual emergence of a genuine dialectical psychology, a *Soviet* psychology, based on dialectical materialism and developed throughout in conscious accordance with its principles (1931 to the present).

This conscious attempt to found a psychology whose methodology is that of dialectical materialism (**101**, **111**, **114**), as well as the need to cope with the specific tasks placed before it by the urgent demands of revolutionary reconstruction from Tsarism to socialism, determined in large measure the particular development that dialectic psychology came to exhibit (**137**, p. 21). Teplov, writing as a Soviet psychologist, therefore states that "the history of Soviet psychology is really the history of the Soviet psychologist's mastery of the Marxist-Leninist methodology . . . [and that] each new step forward of Soviet scientific psychology, each of its large-scale achievements, is witness to a new stage in his creative mastery of Marxism" (**117**, p. 9).

It will be seen that the development of dialectic psychology in the Soviet Union was by no means a smooth and uninterrupted affair. It had its confused and muddled periods, as well as periods of clear direction and steady progress; times also when in retrospect it seemed to be diverted far from its set aims and goals. But the eventuating result was a psychology, firmly set on dialectical materialist principles and on these bases, to all appearances, effectively functioning and forging ahead.

THE PRE-MARXIAN PERIOD

General Observations

The first period, dating from the revolution to 1923 inclusive, is one during which psychology in the Soviet Union may be seen as beset by the same well-known difficulties plaguing psychology in Western Europe and America. Attempted solutions of these difficulties tended to follow the same lines as in western Europe and America, and for a number of years, up until 1930–1931, Gestaltist orientations, Freudianism, Stern's personalistic psychology, Adler's individual psychology, and especially American behaviorism were from time to time more or less in evidence (**28**, **127**, **136**).

However, to overcome the crisis in psychology as they saw it, and under the influence of the continuing revolution, Soviet psychologists during this period came to champion the need of a thoroughgoing reconstruction of the philosophical foundations of psychology in the spirit of a genuine Marxism. Although this reconstruction was subsequently considered to have fallen far short of actual accomplishment, the result at that time, nevertheless, was that so-called idealistic psychology soon ceased to exert any real influence. In fact, its extreme speculative and metaphysical wings, represented by such figures as Losskiĭ and Frank, disappeared altogether, and conceptions, which in Europe are associated with Husserl and the phenomenologists and in the Soviet Union with the names of Lopatin, Chelpanov, and others (**7**, Ch. 8; **137**), were surrendered in favor of the doctrine of "scientific materialism" (**65, 93**).

Instead, developments patterned after American behaviorism came strongly to the fore, especially in the guise of Bekhterev's reflexology. At the same time, also, Pavlov (**79**), the successor to the work of the physiologist Sechenov (**12, 39, 49**), was given, through the personal intercession of Lenin, every opportunity to proceed with his task of laying the foundations for a neurologically based psychology despite the disruptions occasioned by the Civil War[3] (**80**).

The Contributions of Blonskiĭ and Kornilov

In 1921 Blonskiĭ[4] in his book, *Essay on Scientific Psychology* (**21**), entered the revolutionary lists against the traditional psychologies, which he claimed reeked of idealism, and issued the first summons in behalf of a psychology reared on materialist bases. Although he was later criticized for his adherence to American behaviorism, his gross mechanism, and lack of dialectical materialist mastery, his attacks on "university psychology" were considered telling. Highly regarded also was his thesis of the social and class nature of psychological man. "In a class society 'man in general'," wrote Blonskiĭ, "is an empty abstraction; for man's social behavior is determined by the behavior of his class" (**117**, pp. 10–11).

In 1922 Kornilov, in his book, *The Theory of Human Reactions*[5] (**42**), joined the fray against idealism and philosophy in psychology. Although a pupil of Chelpanov, he insisted on the complete separation of psychology from philosophy, both as regards method and object of study, so that it be included in the ranks of the natural sciences. Kornilov, in addition, championed an end to the academic cloistering of experimental psychology, demanding that it not shun to deal with psychological questions pertinent to life and the demands of "socialist construction."

---

[3] The course of events which led to the disappearance of pre-revolutionary Russian psychology and the emergence of "materialist" schools was to be expected, given the thoroughgoing character of the Russian revolution and its planned and unplanned impact affecting the whole of Russian society.

[4] Blonskiĭ was a pupil of Chelpanov (**28a**), who was, along with Nechaevy, N. N. Lange, and Bekhterev, one of the foremost representatives of experimental psychology in pre-revolutionary Russia (**7**, Ch. 6). Of this group Bekhterev is especially to be noted for his program of "objective psychology" which he promulgated in a book (**16**) of like name, published in 1907, six years before Watson's first published espousal of similar ideas (**135**).

[5] This work deals with the measurement of the energetic side of behavior and its energy-expenditure at different levels of reaction. Kornilov was, thereby, led to a new classification of behavior based on the rapidity of the subject's reaction and his inclination for a greater or lesser energy-expenditure.

In 1923, Kornilov issued a veritable call to arms in behalf of a Marxist psychology frankly based on dialectical materialism when, before the First All-Russian Psychoneurological Congress held in Moscow, he presented his paper, "Contemporary Psychology and Marxism" (**43**). This paper signaled the first open struggle between the new "materialist" psychology and the old idealist psychology, dubbed Chelpanovian. It led to the latter's swift demise, thus bringing to a close the first stage in the development of a Marxist psychology.

THE TRIAL AND ERROR PERIOD

It was during the second period, dating from 1924 to 1930–1931, that Marxism may be said to have won for itself recognition as the proper methodological basis of psychology in the Soviet Union. But, as we shall see, this recognition was at best theoretical, since, by the end of the decade, its implementation in actual acceptable psychological theory was judged not to have been satisfactorily accomplished. However, several psychological schools did spring up, each claiming to be the true bearer of Marxism in psychology, each subsequently to be denied its claim.

Bekhterev's School and Contemporary Criticism

During the first decade of this century Bekhterev initiated the laboratory study of "associated reflexes," using dogs and then human beings as subjects for special training. His method of "associated reflexes" is distinguished from that of "condition*al* reflexes" in that he made use of electrical skin irritation as the "chief stimulus" (Pavlov's "unconditioned stimulus"). This stimulus, when applied, provokes a defensive movement. When combined with some originally indifferent stimulus, an "associated reflex" (Pavlov's "conditional reflex") is developed in that the previously neutral stimulus can itself now evoke the same defensive movement (**19**).

For the many who insisted that the chief and direct object of "reflexological" study was the human personality, itself, there were many advantages seen associated with the selection of the motor sphere instead of the secretory, as in the Pavlovian school. Shnirman (**106**), for example, listed the following: the facilitation of transfer of experimental results when human beings are involved; the inclusion of voluntary acts for study purposes; the possibility for an accounting of the effect of antagonists, that is to say, an accounting of not only the positive, but also the negative values of the reflex. In their enthusiasm the partisans of Bekhterev's approach did not hesitate to claim for it priority and superiority to that of Pavlov's.

Pavlov consistently and intentionally operated within the framework of physiology in his classic investigations on the conditioned reflex. Bekhterev, however, went to such lengths as to propose an independent discipline, reflexology, distinct from the physiological study of reflexes and from psychology which it was to replace. Bekhterev affirmed the inaccessibility of the psyche to knowledge and, stressing instead "outer behavior," reduced conscious behavior to combinations of reflexes, essentially no different from those of animals. This stand put Bekhterev in the forefront of the movement in the direction of what afterwards came to be termed "vulgar mechanism" (**93**).

Maintaining that the whole of psychology was rank idealism, inasmuch as psychology made the psyche the subject-matter of its study, reflexology from about 1920 to 1926 acquired tremendous popularity and spread to the point of actually pushing psychology out of school curricula. In the words of Rubinshteĭn, "vulgar mechanistic materialism, relying on the support of the reigning 'methodologists' of those years, triumphed, celebrating the feast of victory" (**91**, p. 68).

Thus, according to Bekhterev's critics (**81**), in reflexology all behavior was reduced mechanistically to elementary reactions of the organism (in other words, to reflexes) so that by their proper combinations all properties and special characteristics of human activity, both psychic and social, could presumably be accounted for.[6] But, in addition to this, in the view of subsequent criticism, Bekhterev, in formulating the philosophical bases of reflexology, unwittingly committeed himself to an idealism as "vulgar" as his mechanism in that he reduced literally everything to "energy."

In the hostile eyes of his critics, out of Bekhterev's characteristic reduction of matter and the psyche to energy there inevitably issued two unfortunate consequences: (1) the mechanical extrapolation of

[6] To accomplish this, besides general reflexology, there were established subdivisions such as genetic, pedagogical, pathological, and collective reflexology. However, Bekhterev, himself, considered reflexology a "biosocial" and "bioconscious" discipline (**19**, **106**).

physical laws over not only to psychic but also to social phenomena; (2) the energizing of the psyche in such a way that the psyche, which initially was denied in its specificity for man, came to be viewed as spread in the form of energy over the whole world—a notion that was denounced as partaking of the nature of an extreme panpsychism (**91**, p. 68). Yet, despite the extent and violence of this criticism, Shnirman was still able to claim in behalf of Bekhterev, one year before the "storm" which relegated reflexology into the limbo of discredited theories broke in 1930–1931, that Bekhterev "stands firmly on the basis of dialectical materialism" (**105**, p. 237).

Kornilov's School and Contemporary Criticism

The attack which Kornilov led on Chelpanovian psychology before the First All-Russian Psychoneurological Congress swiftly became the occasion of a three-cornered argument when the advocates of Bekhterev's reflexology got to propounding zealously their views and assaulting bitterly those of their opponents—Kornilov and Chelpanov. Kornilov's adherents regarded Bekhterev's reflexology as "vulgar mechanism" fast degenerating into direct idealism; whereas the partisans of Bekhterev viewed the position of Kornilov's following as metaphysically tainted in that the latter was still interested in "psychology," which to the proponents of reflexology was *prima facie* evidence of the essentially idealistic character of the thinking of Kornilov and his group (**93**).

The representatives of reflexology were most active in the propagation of their views and, at the Second All-Russian Psychoneurological Congress held in 1924 in Leningrad, they scored their greatest public success (**117**). However, despite the reflexological craze, toward the close of 1923, when a radical reorganization of the Moscow Psychological Institute was pushed through, it was Kornilov who became the director of its new basic staff of young "progressively" oriented researchers.

Kornilov in a series of programmatic papers and articles from the years 1923 to 1925 developed what he conceived to be the most important propositions which must fundamentally underlie a truly Marxist psychology (**45, 117**). He made the following points:

1. Marxism eschews the dualism of mind and matter. Consequently, the psyche is not something opposite to matter, but only a property of "most highly organized matter." Hence, the proposition of materialistic monism is necessitated.

2. However, materialism is subject to two interpretations. One is that of naive materialism (sometimes designated as natural scientific, mechanistic, or

vulgar materialism); the other is that of dialectical materialism. Naive materialism frequently reverts to idealism, and each denial of psychical states as such or their identification with physiological processes is in direct contradiction to dialectical materialism.

3. Furthermore, the essence of man is the aggregate of his social relationships, and the psychology of the individual is impossible that does not deal with the psychology of his class.

4. Also to be kept in mind are the basic principles of dialectics: (a) continuous change, (b) the universal interconnection of phenomena, and (c) the "leaping" or saltatory emergent character of development.

Kornilov's basic points, however correct they may have been in verbal enunciation, were subsequently considered as inadequate to the task of setting up a truly Marxist psychology and insufficiently understood by Kornilov himself; ostensibly as evidenced by his further elaboration of them and the use to which they were put (**11**, **45a**, **93**). Nevertheless, they brought about the complete eviction of Chelpanovian psychology from the scene and, from 1925 on, the steady decline in enthusiasm for reflexology.

As his version of a Marxist psychology, Kornilov proposed "reactology" (**45**, pp. 268–278), that is, "the study of man's reactions with which he responds to outer stimulation" (**117**). The chief method of research was to be the study of the speed and strength of man's movements in his response to earlier conditioned signals. However, it soon became evident that between the grandiose aims of the contemplated Marxian psychology and the scanty program of its concrete content there was fundamental contradiction. The latter could not possibly lead to the former.

Therefore, according to his many critics, Kornilov never did really manage more than a programmatic declaration of principles. They felt that "reactology," the system of psychology which Kornilov devised to meet the demands of the thoroughgoing Marxism that he contemplated, was also, in the last analysis, after all only an "eclectic mechanistic formulation based on psychophysical parallelism" (**37**, **93**), and thus could hardly be said to have fulfilled the program he had set for himself and psychology in general. As a matter of fact, the *Great Soviet Encyclopedia* goes so far as to accuse Kornilov of "not having [at that time sufficiently] mastered the Marxist-Leninist methodology," so that, in its opinion, it was not surprising that his "reactological psychology, under the strong influence of Wundtian doctrine, was itself a victim of the coarsest mechanism interwoven with idealistic error" (**45a**, p. 323).

Although in the 1924–1925 methodological articles of Kornilov one finds Marxist-Leninist formulations like "consciousness as a reflection of existence" and "perception as the subjective reflection of existence, whose reality is independent of consciousness," these gave way in 1926, when his *Textbook of Psychology* (**44**) appeared, to a formulation which was criticized as deeply in error and anti-Leninist; namely, "the psyche and consciousness are the subjective reflection of physiological processes." In the last analysis, then, in reactology the psyche and consciousness, being mere duplications of physiological processes, could not be the true subject-matter of psychology, but only behavior.

Many were dissatisfied with the banishment of consciousness from scientific psychology. There arose, therefore, a "struggle in behalf of consciousness." As far back as 1925, when the first collection of research conducted at the Moscow Psychological Institute was published, Vygotskiĭ protested that, "in ignoring the problem of consciousness, psychology, itself, cuts off access to the investigation of those problems of the psychology of behavior that are to any degree complex" (**117**, p. 15; **129**). The problem of consciousness, however, had to wait until the next decade before its acceptable resolution. Reflexology would have none of it; reactology could not cope with it.

The reactological views of Kornilov were undermined not only by critical Marxian analyses, as already noted, but also in part by the experimental investigation, begun by Vygotskiĭ and continued by his co-worker, Solov'ëvyĭ, in which it was shown that a "dominant tendency" of behavior, expressed in the form of a "conscious aim," can subordinate to itself unconditioned reflexes "infinitely surpassing it in strength." Thus, if a person is required not simply to perform certain actions, but to perform them in spite of interference on the part of stimuli irrelevant to the task, then these stimuli are psychologically no longer distractions, but obstacles to be overcome, and the rate of basic work under such circumstances may exceed that under quiet conditions (**117**, p. 14).

This revealed the possibility of breaking past the limitations imposed by the reactological views of man as a merely reactive being, that is to say, a being passively respondent to stimuli. Even reactologists were not long in beginning to emphasize the directed activeness of man.

It should be noted that throughout Kornilov's reactology the sharp and militant mechanism of Bekhterev's behaviorism became more compromising and eclectic in character. This was because reactology essentially attempted a synthesis of the two extremes—behaviorism and introspectionism, professedly in strict accordance with the principles of dialectical materialism (**45**, **93**).

As a dialectical materialist, Kornilov thought it was proper to regard the subjective correlates of physiological processes taking place within the nervous system as the introspective aspect of material phenomena.

In this he was affirming the principle of the unity, but not the identity, of the physical and the psychical (**106**).

Kornilov came then to believe in the possibility of establishing a demonstrable causal coherency between physical phenomena; but between physical and psychical, or between psychical and psychical, he averred its impossibility. Thus, he saw the "objective" method as the only scientific method. The method of introspection,[7] though by no means ruled out, could offer only subjective description, which without concomitant control through objective data, was conceived as having nothing to offer of scientific value. In other words, psychology was, after all, the science of behavior; but behavior, itself, in the last analysis must be regarded as a combination of reactions responsive to external stimuli.

Kornilov viewed these reactions as energy processes and strove to relate them to inner psychic phenomena. However, his critics regarded his efforts as smacking too much of idealistic psychophysical parallelism and as too mechanistic in conception. This theoretical performance of Kornilov's, therefore, could not be successfully carried off. Kornilov, despite good intentions, was regarded as having failed, along with Bekhterev, to establish a "truly dialectical psychology—a psychology adequate to the demands of "socialist construction"—and "after sharp discussion in 1931 he acknowledged his former errors" (**45a**, p. 323).

The Influence of Pavlov and Ukhtomskiĭ

In 1923 Soviet psychologists were confronted with the challenging task of incorporating and utilizing the brilliant discoveries of Pavlov, presented in his book, *Twenty Years of Objective Experimental Study of the Higher Nervous Activity of Animals* (**82**). This was ably performed in Kravkov's *Essay on Psychology* (**51**) and Dobrynin's *Introduction to Psychology* (**29**). In addition, Vygotskiĭ, basing himself on the Pavlovian theory of the conditioned reflex, developed during the years of 1925–1926 the ideas of Sechenov[8] on the reflexive character of the psyche (**7**, Ch. 5; **39**).

Although Pavlov's theory was viewed as providing at last a physiological explanation for the various laws of association of the 19th century, the explanatory reduction of all psychic phenomena to this one theory has been since frowned upon and regarded as incorrect in principle. Furthermore, the Pavlovian theory has been affirmed as inadequate for even an understanding of the physiological bases of human behavior, let alone its psychic aspects (**1**; **30**; **34**; **36**; **80**; **91**, pp. 462, 500).

---

[7] According to Kornilov, the materialist cannot acknowledge non-material phenomena, since for him there exists only one single series of real processes and not two in psychological interaction or parallelism. But the singleness of this series does not exclude the possibility of different non-identical approaches to it, and the introspective approach is one of them.

[8] I. M. Sechenov (1829–1905) is regarded as the father of Russian physiology. His works were attempts to explain all behavior of man as composite reflexes (**49**).

However, Pavlov's immense prestige and authority induced too credulous and uncritical an attitude toward physiological work in general. The result of this hyper-enthusiasm for anything physiological was a rash of rather impetuous generalizations and an attempt, on the part of physiologists, to trespass onto other fields and, conversely, a simultaneous eagerness, on the part of the workers in these invaded fields, to "physiologize" or "biologize" their own subject-matter of their own accord. The very existence of psychology as an independent discipline was threatened and, as already noted, reflexology, in particular, did actually manage a partial displacement of psychology (**28**, **93**).

Mention should be made of Ukhtomskiĭ (**122a**) and his "theory of the dominant" (**50**, pp. 33–67; **122**) which exercised a great influence on Soviet psychological thought. From 1923 on, this Soviet physiologist developed in a series of papers and articles the idea of the "dominant" as a seat of temporarily heightened neural excitation, deflecting to itself stimuli from other possible seats and thus dominating and determining the joint action of neural centers. Ukhtomskiĭ was led to this theory by such observations as the following: (1) A cat, surprised in the act of micturition by a dog, seems rooted to the spot, paralyzed not by fear but by its momentary activity; (2) electrical stimulation of the motor cortex of an animal with a full rectum will not lead invariably to movement of the corresponding extremities but frequently to defecation. Only when the rectum has been emptied will electrical stimulation of the motor cortex always evoke the customary motor responses. To many the theory of the dominant seemed to bring nearer to realization a physiological explantion of goal-directed behavior and, according to Dobrynin, that of attention. In any event, the theory was effective in as much as it, at the very least, stimulated much experimental work (**36**, **125**).

Vygotskiĭ's School and Contemporary Criticism

Beginning in 1928 the problem of development attracted considerable interest and was the subject of special attack by a group of young researchers, among them Lurii͡a and Leont'ev, headed by Vygotskiĭ. From their joint efforts there emerged their "theory of cultural historical development,'' in which they strove (1) to make the foundation stone of a Marxian psychology the principle of development, (2) to approach the process of psychic development dialectically (**66**), that is, to stress the qualitative distinctiveness of its several stages, (3) to explain man's psyche as a historical developmental product, and (4) to discover the explanatory principle behind the higher psychic processes such as logical memory, voluntary attention, thought, and speech.

The originators of this theory devised original methods of experimental research on the development of these processes in children and established a series of special laws and facts of a most interesting kind (**58, 64, 77**). Examples of their work are well represented in Leont'ev's monograph, *The Development of Memory* (**57**) and Vygotskiĭ's *Thought and Speech*, published posthumously in 1934 (**132, 134**).

However, the originators of this theory were criticized as having fallen, like Kornilov previously, into the same trap that "characteristically" ensnares the "bourgeois" psychologist: namely, the indefensible practice of creating one's own particular psychology and a corresponding generalized system on the basis of isolated laws and scraps of data derived from one's own special researches.

Vygotskiĭ and Luri͡ia thus tried to show that "cultural development consists essentially in the adoption and assimilation of those behavioral modes based on 'sign'-utilization and employment as a means for carrying out this or that operation" (**117**, p. 16). The proverbial "tying a knot so as not to forget" became in its own way for them a symbol of each and every "cultural" operation. There was drawn a sharp division between "natural development" (e.g., the organic development of memory) and "cultural development" (e..g, the mastery of symbolic methods of recall). In addition to the above, the completely abstract and formal approach of this theory to the problem of the historical development of the psyche came under fire—an approach which was criticized as overlooking the concrete course of social development, the class struggle, etc. Also censured was its overly abstract approach to child thought with its consequent neglect of his concrete social environment, his concrete life and activity (**91**, pp. 338–339).

Vygotskiĭ ahd Luri͡ia were taken to task on yet another score. They were accused of exhibiting a failing in common with that of many other Soviet psychologists of that period—the uncritical borrowing of all manner of theories from the "bourgeois" psychologies—a practice which by 1929 was very much in evidence. This practice was, indeed, a fact not to be glossed over, however insistent were the distinctions finely drawn between Soviet and "bourgeois" psychologies of kindred spirit. Blonskiĭ,[9] for example, was a behaviorist of Watsonian derivation, however much he opposed Soviet to American behaviorism, as when he remarked in 1925 that the latter "never would dare and hence could not give a correct explanation of American bourgeois behavior and disclose [for instance] the true essence and motives of American senatorial behavior" (**117**, p. 17).

Also placed under especial attack was the wide-spread opinion, expressed specifically by Vygotskiĭ, that (1) it was necessary to distinguish in foreign psychological schools two aspects: (a) "the factual basis" and (b) "the theory constructed thereon"; and (2) the denial of the second need not necessarily lead to rejection of the first. It was pointed out by critics that even among men of the greatest integrity "the factual basis" has its content in effect determined by the "theory." These critics maintained that the psychologist can always "extract" from out the manifold welter of reality precisely those psychological "facts" which his theory demands and, further, that even among individual "bourgeois" psychologists this "factual basis" has often been looked upon as erroneous and displaying more the essence of prejudicing interpretation by a given school than the essence of neutral fact (**117**). This criticism was especially damning in that psychological theories were and even now are considered as having philosophical bias and political leaning, so that "bourgeois" psychologies are *ipso facto* reactionary in essential character.

General Observations

During the second period there were, as we have seen, several ambitious attempts at the construction of general theories, conceptions, and systems—often under the banner of an aggressive mechanism, but all judged unsuccessful from the dialectical materialist point of view. At the same time a significant increase in the range of psychological work took place, particularly in child, pedagogical, and industrial psychology. An independent journal, *Psikhologii͡a*, was established for psychology, many collections of experimental work were published, and a special section accorded it in the philosophical faculty of higher educational institutions.

---

[9] Blonskiĭ was one of the introducers of behaviorism into Soviet psychology. Nevertheless, Blonskiĭ valued the social determination of behavior and shunned the extreme mechanistic point of view of the reflexologists. His own work displays a distinctly genetic character. As a matter of fact, he regarded psychology as a branch of genetic biology in which consciousness is equated to the "vigilant" life. According to Blonskiĭ psychology can best be thought of as the science of the organism's "sleeping" and "vigilant" life, all of whose aspects coexist at their various phylogenetic levels within the same organism in integral and dynamic interrelationship.

Many books of a generalizing character appeared; for example, Kornilov's *Textbook of Psychology, Expounded from the Point of View of Dialectical Materialism* (**44**), Vygotskiĭ's *Pedagogical Psychology* (**130**), Basov's *Methods of Psychological Observation on Children* (**15**)[10], Blonskiĭ's *Psychological Essays* (**23**)[11], and Dobrynin's *Introduction to Psychology* (**29**).

Despite the great preoccupation with dialectical materialist principles by Soviet psychologists, only one "materialist" proposition was subsequently considered as rightly taken—*the psyche is a property of highly organized matter*. But the recognition of the psyche as a property of the brain was in itself insufficient for the establishment of a Marxist psychology in as much as there was no real idea of what this property should consist. As will be later seen, Lenin's "theory of reflection" was to provide the clue for the dialectical psychologist (**55, 56**).

Meanwhile, toward the end of the second period, in 1929 and 1930 the center of research at the Moscow Psychological Institute was focussed on the problems of social and collective psychology,[12] and a great deal of attention was thereto directed, not with much success, however, from the standpoint of subsequent criticism. Kornilov headed the section of general psychology, Rybnikov—child psychology, Borovskiĭ—zoöpsychology (habit formation in animals), and Shpilreĭn—psychotechnic ("professiography," fatigue, industrial training, etc.) (**28, 93**).

Toward the end of the 1920's enthusiasm for various "bourgeois" psychological schools had become strongly pronounced. Soviet psychologists discerned affinities among these schools which made them seem "close" or "related" to Soviet psychology. Especial enthusiasm was reserved for German "holistic psychology," and, in particular, its Berlin branch, Gestalt psychology. However, criticism was unceasing, and it was driven home that the declarations of the Gestaltist school against mechanism and the "alluring vitality" of its experimental methodology could serve as a screen for its "idealistic" essence only among those Soviet psychologists whose deep philosophical and political convictions had been blunted. For example, Borovskiĭ in 1929 wrote:

---

[10] Basov held as primary for the child psychologist objective registration and analysis of those data which child life offers us. He maintained that an understanding of the fundamental laws of child behavior can come only from a correct understanding of single acts of behavior and their causes. This, he said, could be achieved only through the method of objective and systematic observation.

[11] In this book Blonskiĭ tends to regard man's psyche dynamically as a gradual development of biological processes in which there coexist in him different processes belonging to different developmental states.

[12] In collective and class psychology the subject of study is never viewed as "man in general," but always as man, the representative of a social class.

> Some of our junior fellows are showing considerable interest lately in the German Gestalt psychology . . . . Certain principles in it [seem to] conform to fundamental dialectical postulates. [But] the subjectivism and idealism, which are very prominent in Gestalt psychology, cannot . . . be approved of . . . . [As for American behaviorism], we shall never agree with a behaviorism trying to be purely empirical (for "no philosophy means bad philosophy") or with a behaviorism tending to eclectics . . . . Behaviorism [needs rebuilding] on the philosophical basis of dialectical materialism (**28**, p. 184).

These psychologists were chided for seeking "fellow-travelers" among bourgeois psychological schools instead of devoting their energies toward the construction of a truly Marxist psychology.

Among the positive achievements of the second period the development of particular experimental methodologies must be included—that which even subsequent criticism concedes. To be singled out for special mention are Luriia͡'s method for the study of emotional states (**68**, **69**, **70**, **72**, **76**); the methods for the study of memory, thought, and speech, devised by Vygotskiĭ's students: Sakharov (**99**), Leont'ev (**57**), Zankov, and others; Basov's observational method with preschool children (**15**, **106**); and Rybnikov's "biographical method" involving the use of diaries, autobiographies, etc.

Crisis, "*Housecleaning*," *and Reorientation*

Despite its accomplishments, however, toward the end of the second period psychology suffered great loss in prestige and sacrificed the greater part of ground previously gained. There were several reasons for this retrogression and the chief ones were perhaps bound up with psychology itself; for the dominance in it of certain mechanistic tendencies are now seen as having entailed two unfortunate consequences: one which led directly to the liquidation of psychology itself as a science (*via* reflexology) and the other which led indirectly to the same result by bringing about in psychology a state of general sterility (**93**).

The second period in the development of psychology in the Soviet Union may be viewed as terminating in 1931, at which time there was a culmination in the sharp crescendo of discussions directed against the dominant mechanistic tendencies rooted in the then reigning psychologies: reflexology, reactology, and related conceptions. The result was their demise; for even where reflexology and related views seemed most vigorously to do battle with the "idealistic" psychologies in behalf of a Marxist psychology, these were subsequently perceived as mere struggles between schools within the framework of traditional "bourgeois" psychology.

This development in Soviet psychology cannot be understood apart from those developments undergoing acceleration within the larger framework of Soviet society itself. In a very real sense 1929 must, in fact, be regarded as the "year of the great divide" (**113**, p. 264) for psychology as for the rest of the Soviet Union, because in this year decisions were made by Stalin and the Communist Party which were to have long enduring effects throughout the whole of Soviet society. Prior to 1929 one might say that sundry "capitalistic" elements and influences had been more or less tolerated; subsequently they were to be studiously, and in many cases ruthlessly, weeded out—all this as the passionate accompaniment of the adoption of the famous five year plan calling for the establishment of heavy industry and the collectivization of farms in the Soviet Union.

The struggle against "reactionary" influences was not long, therefore, in finding expression on the philosophical "front" and led to a busy reinvestigation, under Marxist-Leninist principles, of basic and derived ideology. As a matter of fact, throughout 1930 there took place among Soviet philosophers much discussion whose aim, at Stalin's suggestion, was two-fold: (1) the liquidation of "anti-Marxian" idealism, on the one hand, and vulgar mechanism, on the other; and (2) the development and exploitation of the philosophical heritage bequeathed by Lenin (**117**).

As on the philosophical front, so on the psychological front. From 1930 to 1932 much discussion on an organized scale was in evidence. The discussions on reactology, for instance, were initiated by the communist cell of the Moscow Psychological Institute. These particular discussions, incidentally, held during the first half of 1931, turned out to be crucial for Soviet psychology. For this reason 1931 may be rightly viewed as terminating somewhat emphatically the second period of development in dialectic psychology (**93**).

The public acknowledgment by Kornilov of his "errors" has already been alluded to. This took place during the reactological discussions mentioned above. Incidentally, we here see illustrated in this renunciation of reactology by Kornilov, which superficially might be equated to recantation, a practice by no means confined to Soviet psychological circles; one that is assiduously cultivated and in its extent almost unique to the Soviet Union—the practice of public criticism and *samokritika* (self-criticism) in forum, sessions, open and closed, and in print (**99a**, **137**). For example, psychological gatherings are even now convened whose primary agenda is discussion and evaluation of newly-published books, theories, programs of action, etc. (**41**, **47**, **104**, **117**, **120**).

In this connection Zhdanov, as a matter of fact, was not far off when he maintained that it was precisely out of such discussions over the years and particularly during certain periods (which Soviet psychologists have since regarded as critical for the formation of a dialectical psychology) that there were hammered out the most important of the propositions deemed essential for the establishment of a truly Marxist psychology, that is to say, a psychology securely founded on the principles of dialectical materialism, as first formulated by Marx and reworked by Lenin. The history of dialectic psychology is one in which criticism and *samokritika* have played no mean roles (**117**, **137**).

Two insistencies underlay all these discussions. Baldly stated, they were a call for (1) a heightened sense of "Bolshevik party-spirit" in Soviet psychology and (2) a Soviet psychology applicable to and commensurate with the urgent demands of "socialist construction." Accordingly, there arose a sharp reversal in the attitude of Soviet psychologists to foreign psychological thought. Before, many were prone to look on foreign psychologies as somehow supplying the "last word in psychological thinking" or, at least, as worthy of serious study for their "factual bases" despite the untenability, from the Soviet point of view, of superimposed theories. Now it was declared evident that to follow the lead of any of the major schools of foreign psychology such as behaviorism, Gestalt psychology, etc., was to get one's self up a blind alley from which there was no escaping (**117**).

At the same time the point was driven home that, if psychology shunned the pressing problems of real life and applied practice, whatever the claims may be that it might make for itself, it could not claim recognition as a science geared to Marxism-Leninism. Only in the course of solving those problems generated out of the very demands of everyday living and "socialist construction" was it possible, so it was averred, to establish in Soviet psychology a truly scientific theory.

In the fundamental reorientation that took place in Soviet psychology, Lenin's basic philosophy was decisive. Two works of Lenin's must be singled out as having exerted an especial influence. They are his *Materialism and Empirio-Criticism* (**54**) and his *Philosophical Notebooks* (**55**, **56**), the latter first published posthumously in 1929–1930. Discussions based on the material of these two books in particular led to a reexamination of basic methodological problems in psychology and to the "conviction," derived from the study of the latter work, that without thoroughgoing acceptance and mastery of the Leninist "theory of reflection" the construction of a Marxian psychology would be impossible (**93**, **117**).

The Leninist conception of the psyche and consciousness as the reflection of an objective world, it was believed, established at long last the means for "materialistically" substantiating the notion of psychology as the science of man's psyche in aggressive opposition to the notion of psychology as the science of man's behavior, understood as an aggregate of outer reactions. Consequently, behaviorism in all its forms—reactology, reflexology, etc.—simply folded up under the vigor of the concerted attack to which it was subjected.

Not only was behaviorism decried as leading to an inadequate psychology, but its "reactionary," "political" character was held up for excoriation. In behaviorism's "renunciation of consciousness" and "approach to man as a reacting automaton" were discerned the not too subtle influence of capitalism; so that in 1931 Vedenov, summarizing in part the discussions on reactology, could declare:

> The whole reactionary character of [behaviorism's] approach to man is quite in the open. Behold the ideal of capitalism—man, the automaton, a robot, whose action one can compel as one wills! Behold the dream of capitalists the world over—a working class without consciousness, without mentality, whose actions are supinely subject to the training whims of the exploiter! That is why it is in America, in that bulwark of contemporary capitalism, that this theory of the man-machine was so powerfully developed and so stubbornly maintained (**117**, p. 21).

As a consequence, not only did consciousness and its directive and regulative roles in man's behavior become questions of paramount urgency, but the very approach to the investigation of a majority of psychological problems was radically changed—an approach first systematically exhibited in researches on habit formation. From this time on, a new conception of habits began to emerge—a conception which saw in habit formation in man a process essentially conscious. The question, "What constitutes the role of consciousness in the process of habit-formation?" came to pose the essential problem for the Soviet investigator in this area of psychological study (**91**, ch. 16).

The psychological discussions, whose consequences we have been indicating, set the final seal on the fate of reflexology. Even prior to this time, things had not been going too well in the reflexological camp. Bekhterev had died in 1927 (**107**), and shortly thereafter, in the course of lively sessions marked by much criticism and self-criticism, a breach sundered the hitherto solid phalanx of his following. One group tread in Bekhterev's footsteps, espoused to a militant mechanism and a compounding of his "errors"; the other, having subjected the methodological foundations of reflexology to critical analysis, sought a new direction.

It was out of this latter group that there arose such active workers in behalf of a dialectical psychology as V. N. Mi͡asishchev and B. G. Anan'ev. Anan'ev's paper, published in the journal, *Psikhologii͡a*, at the termination of the "reactological discussions" and since praised as a model of "Bolshevik self-criticism," set forth so definitive a methodological critique of reflexology that to this day it is considered reflexology's fitting epitaph (**117**).[13]

THE DIALECTIC PERIOD

General Observations (1931–1936)

At the beginning of the third period in 1931 the prestige of psychology had, as already noted, fallen off to a very considerable degree—a state of affairs reflected by a drop in the number of psychological works published. However, it was felt that the liquidational tendencies of "vulgar mechanistic materialism" in relation to psychology were in principle, at least, overcome, and faith in the possibilities of a truly dialectical psychology sought affirmation in the wide world of practice. New centers for psychological work were established throughout the Soviet Union; e.g., in Georgia, in the Ukraine (Kharkov, Kiev, and Odessa), and in Leningrad (Gert͡sen [Herzen] State Pedagogical Institute and the Leningrad Brain Institute). Moreover, with the ambitious programs of industrial and educational expansion zealously undertaken by the Soviet Union, interest, already pronounced, in child, pedagogical, and industrial psychology continued steadily on the rise (**93**).

Vygotskiĭ's theory of the cultural development of man's higher psychic functions continued under discussion. The basic tendency of Vygotskiĭ's theory, whose ideas were often in close coincidence with those of the French sociological school (as represented by Blondel (**20**), for example) was to forward in psychology the idea of development and the principle of historicism. However, as previously pointed out, the theory was adjudged defective and interest in it gradually diminished.

The challenge to comprehend the human psyche in the light of Lenin's theory of reflection pushed to the fore the problem of sensation and perception as constituting the initial step toward knowledge of the world. Previously, hardly any attention had been directed to this particular problem in Soviet psychology. The reverse now became true. The sensations, especially the visual, were subject to intensive research. Out of this research there emerged, in the work of Kravkov (**53**) and

---

[13] Translations from the Russian into English are hard pressed to keep pace with the kaleidoscopic developments in the Soviet Union and unintentionally contribute to much erroneous misconception as to the past and present state of affairs in the Soviet psychological world. Thus, the reflexological and reactological articles by Shnirman (**105**) and Kornilov (**45**) respectively were already obsolete at their date of published translation in 1930, and the translators of Bekhterev's *General Principles of Human Reflexology* could make in 1932(!) the incredible statement that "reflexology is the dominant note in Russian psychology today, and the late Bekhterev is the guiding light of humanistic science in the U.S.S.R." (**19**, p. 15).

Lebedinskiĭ among others, the theory of the "interaction of sensations" (**8**, pp. 259–260). This theory presented the thesis that by means of simultaneous excitation or induction from other sense organs the threshold of sensation can be lowered for a given sense organ (**91**, p. 156).

In 1932 and 1936 there appeared the first and second editions of Kravkov's book *The Eye and Its Functions* (**52**), lauded as one of the best generalizing pieces of literature on the psychophysiology of vision. However, not only were the theoretical aspects of visual sensation and perception pursued by researchers, but also their practical or applied aspects; for example, the work of Shvart͡s and Artemov on typography; work in the field of military camouflage, and so forth (**117**).

There was also considerable research done on memory and thought processes. Besides the earlier indicated work of Vygotskiĭ and Leont'ev, to be singled out for mention are the books of Blonskiĭ, who had by now turned aside from pedology, an "anathematized" development of pedagogical and child psychology shortly to be discussed. They are *Memory and Thought* (**35**) and *The Development of Thought in the School Child* (**26**), both published in 1935.

In 1934 there appeared an article by S. L. Rubinshteĭn titled, *The Problems of Psychology in the Works of Karl Marx* (**88**), which was destined to exert considerable influence on the development of a dialectical psychology. In this article Rubinshteĭn subjected to careful analysis those statements and pronouncements of Marx relevant to a dialectical psychology, and his analysis of Marx's ideas on demands and abilities is regarded even now as especially significant (**117**, **119**).

On the basis of this analysis Rubinshteĭn proposed a new approach to the solution of such basic problems of psychology as those of development, personality, and consciousness in its relation to activity. The essence of this new approach is strongly suggested in Rubinshteĭn's fundamental psychological proposition—that of the "formation of the psyche in the process of activity"—a thesis which was to be developed in considerable detail in his book, *The Bases of Psychology* (**90**), which appeared in 1935, and in his later works (**91**, **97**).

In the field of pedagogical psychology there was much stirring (**31**, **32**, **84**, **108**, **116**). In this connection the 1931 decree of the Central Communist Committee on the primary and secondary schools (**84a**) must be alluded to. This decree sought to suggest the scope and direction of pedagogical psychology and is the forerunner of the famous 1936 decree on "pedological distortions" (**84b**) in the Soviet school system—the decree which put an end to pedology once and for all.

The 1931 decree ordered the pedagogical research institutes to concentrate their efforts on the study and generalization of experience and data gained by the "practical" workers in the schools. In other words, "bourgeois" pedagogical psychology was no longer to be slavishly aped. "Bougeois" pedagogical psychology was seen as too much the lackey of theoretical psychology; as too busied the applier of the latter's conclusions, derived, not from the concrete pedagogical situation, but from the remote experimental laboratory. This relationship was now to be reversed. The psychological laws and properties of pedagogical situations were to be discovered and worked out in the very course of the pedagogical process itself and not apart from it. An illustration of this new approach is Anan'ev's monograph, *The Psychology of Pedagogical Evaluation* (**10**), published in 1935.

Pedology

During the second half of the 1920's a special discipline, pedology (**32**), got under way, having gathered momentum with the publication in 1925 of Blonskiĭ's *Pedology* (**24**). This book played an important role in the development of this special discipline. In it pedology is no longer looked on as mere child psychology or as experimental pedagogy, but as a genetic science dedicated to the study of the growth of the child.[14]

Interest in pedology mushroomed and by January, 1928 the First Russian Congress of Pedologists had been held. Many prominent psychologists became interested in this new field, from among whom Blonskiĭ, Basov, and Vygotskiĭ are to be singled out.

The psychological discussions of 1930–1931, which today are seen as heralding Soviet psychology as a single science whose various subdivisions are unified by their common methodology and common problems, strangely enough did not materially hamper the growth and spread of pedology. Despite increasingly unfriendly comment, it continued to flourish and managed somehow to bypass the fate accorded the other alledgedly mechanistic psychologies, that is, until 1936 when it was singled out for official censure in a special decree (**84b**).

Pedology came to be essentially an adaptation and development of ideas drawn from western educational psychology with its familiar paraphernalia of measuring techniques, tests, and so forth, and has been described, especially since 1936, as a "vicious pseudo-scientific bourgeois importation" (**32**). This special discipline is still regarded as having wrought incalculable harm to the Soviet school system, and the work of the aforementioned psychologists is represented as precious years lost to worthwhile psychological research.[15]

---

[14] Luriî͡a in 1928 referred to pedology as "the science of the growing organism and its symptom complexes (including the somatic, psychological, and social peculiarities of the child)" (**65**, pp. 347–348).

It must not, however, be concluded that pedology fell into any sudden disfavor and that because of official ukase. Throughout the years of its growth and spread in the Soviet Union pedology had been subject to unceasing attack by many psychologists particularly for its basis in "testology" which is in itself viewed as a pseudo-science. In books and in the journal, *Psikhologii͡a*, there are to be found not a little criticism of test-methodology. Kornilov, for example, in 1928 wrote that "in absence of solid preliminary substantiation the method of tests degenerates into [an activity for which scientific character is claimed, but which is really] a game with jackstraws. . . ." (**117**, p. 19).

The 1936 decree on pedological distortions in the Soviet school system accused pedology of having monstrously perverted the educational process through its regarding the development of the child and the growth of the personality as basically processes fatalistically predetermined by heredity and the elemental action of a depersonalized environment. The decree, furthermore, discerned a "reactionary" character in the principal laws of pedology, since (1) these "laws" unjustly presumed the fatalistic dependence of a child's destiny on biologic and social factors, that is to say, on the influence of a blind heredity and some sort of "irrevocable environment" and (2) these "laws" were shown to be in "howling contradiction to Marxism and to the whole practice of socialist construction, which was [seen as] successfully transforming people in the spirit of socialism and rooting out the remnants of capitalism in the economy and consciousness of men" (**84b**).

In addition, pedology's chief method of investigation—that of "testology"—was pronounced a "downright mockery and insult against school children" and "in opposition to both the real aims of the Soviet school and healthy good sense" (**84b**). Therefore, although the aim of pedology was ostensibly a thorough and many-sided investigation of the child, it was accused of having, in practice, degenerated into "investigation" of children by "standardized" tests. The position was vehemently maintained that such tests may outwardly lay claim to scientific character through their pretense of ability to "measure" intellectual endowments, character, hereditary potentialities, etc., but that in reality they have been perverted to the service of a pseudo-science and, in principle, differ little from graphology and chiromancy. This view is still the current one and is strongly upheld (**91**, pp. 34–35).

---

[15] Psychotechnic (**33**, **102**, **112**, **126**), which developed contemporaneously with pedology and which became a special form of vocational and industrial psychology, likewise was denounced as a pseudo-scientific bourgeois importation and eventually banned along with it.

Pedology, then, despite Marxist trappings, was denounced as anti-Marxist and evilly akin to American "testology" which is viewed as serving, to all intents and purposes, the reactionary interests of capitalism in that "tests happen to be an exceptionally convenient weapon with whose assistance it is possible, with seeemingly complete objectivity and impartial operation with numbers, to bring about the desired selection of workers, to 'demonstrate' the psychic inferiority of the unemployed, to confirm the special talented nature of the children of the exploiting classes, and so forth" (**117**, p. 6).

For similar reasons "psychotechnic" was denounced. Initially psychotechnic was designed to perform the functions of all manner of applied psychology. Subsequently, by psychotechnic was understood particularly that branch of applied psychology which was interested mainly in problems of rationalization of work and vocational selection. It, too, made great use of standardized tests and suffered the same fate as pedology at about the same time.

DIALECTIC PSYCHOLOGY

General Aspects

The liquidation of the "pseudo-sciences" of pedology and psychotechnic served to close the gap between theoretical and applied psychology and facilitated the spread of psychology into fields from which it had previously in large measure been debarred for the reason that these fields had been preëmpted by the pedologist and the psychotechnician. The fact to note is that out of the pedological crisis and due to the increasing progress of psychology toward the goals set for and by the Soviet psychologist, there came to eventuate with the years a flourishing cohesive psychological system which has successfully weathered the exacting years up to the present. Furthermore, on the basis of available evidence, this system seems to have been quite fruitful in both theoretical and experimental investigations as well as in practical applications. The year 1936 may then be regarded as a real turning point in the development of a unified dialectical psychology; one which psychologists in the U.S.S.R. could call a *Soviet* psychology.

However, even before 1936 a definitely dialectical psychology was already noticeably well along the road to satisfactory formulation. Razran, for example, was able to report in 1935 the following:

> Until a very few years ago, dialectical psychology was largely theoretical and philosophical. It debated body-mind problems, sought theses and anti-theses, quoted Marx, Engels, and Lenin. In practice, it borrowed heavily from

> American and German psychology, often little critical whether the borrowed techniques, methods, and even interpretations were in line with its own philosophical principles and social objectives. This has changed now. Dialectical psychology has become a complete, independent, all-embracing system, with a unique selection of problems and appropriate techniques and methods. Its main tenets are materialism, dialectics, psychological levels, and above all an emphasis upon the social rather than the individual aspect of behavior. Indeed, consciousness itself, although admittedly biological in origin,[16] is held to be in reality always social in content and *the heuristic value of the conditioned response is not in interpreting social behavior biologically, but in modifying biological behavior socially* (**86**, pp. 23–24).

Razran's observation, however, that dialectic psychology places "above all an emphasis upon the social rather than the individual aspect of behavior," risks misconception as to the current position, if it is taken to mean the opposition of the social to the personal and the latter's minimization in the psychological scheme. Instead there is in dialectic psychology an insistent emphasis on the "unity of the social and the personal." Therefore, some amplification is in order.

Razran's observation is at present applicable only if it is understood as meaning that the study of individual behavior cannot be undertaken in an ahistorical sociological vacuum, since the Soviet psychologist regards as fundamental the thesis that man, being a member *of* and *in* society, cannot be studied apart from it. And this belief is reflected in his profound distrust for approaches that sacrifice the individual at the altar of broad psychological generalizations which, in the very process of their initial conception and subsequent theoretical formulation, have ignored the concrete activities of concrete individuals in concrete situations (**62**, **63**, **97**). In other words, Soviet psychology claims to be an individualized psychology, but the individualization is that of a *social* being and not of some abstract entity existing in a personal world "locked off" unto itself.

The stress on individuality partly accounts for the hostility of the Soviet psychologist toward that "indiscriminate orgy of standardized tests" which drowns out the individual in a sea of "impersonal statistical averages" (**93**). Pedology and current American practice, as the Soviet psychologist views the latter, are particularly singled out as pernicious examples of a profligate "testology" with its "fatalistic philosophy of static norms ignorant of and detached from the dynamics of mental development" (**85**).

It is true, as Razran has pointed out above, that what passed for dialectical psychology, especially prior to the third decade, did in fact borrow heavily from the "bourgeois" psychologies. Gestaltism, for example, was very well received by Kornilov, who, duly recording his good-will in *Psychologies of 1930* (**45**), cited Gestalt psychology as "thoroughly dialectic" in its exemplification of the Marxian "law of mutual dependence of quality and quantity" (**45**, p. 255).

---

[16] Contemporary Soviet psychology, however, ascribes a social genesis to consciousness.

Certainly, Gestaltism may seem, at least outwardly, to exhibit kinship with dialectical materialism. Gestaltism holds that out of the union of combinatorial parts new qualitites or wholes may emerge. Similarly, dialectical materialism maintains that mere quantitative accretion on one qualitative level can lead emergently to a discontinuous "leap" or *skachók* to another level, qualitatively different from that which furnished the springboard for the saltation into novelty (**111**).

Even Freudianism has been made to exhibit kinship with dialectical materialism (**87**; **91**, p. 67; **124**). Freudianism makes much of the existence of opposing dynamic psychological mechanisms, their struggle, and tentative resolution on a new plane. Dialectical materialism likewise affirms "the unity of opposites" which, through mutual conflict, evokes the emergence of new characteristics (**67**).

Yet the kinship suggested has been adjudged in either case to be at best superficial and rejected (**38**, **87**, **100**). Freudianism is decried as a particularly vicious theory serving as an "idealistic" vehicle for a reactionary decadent irrationalism[17] (**124**). And Gestaltism is criticized for being not only idealistic in its theoretical conception, but also mechanistic in its operation (**35**; **91**, pp. 58–59).

Of the developing dialectical psychology Razran writes:

> The Soviet psychologist's dialectical materialism commits him to the recognition of dynamic opposing tendencies in psychological phenomena, and—what is even more important—to the emergence, through the conflict of these tendencies, of new qualities, emergents, or higher psychological levels. These new emergent levels of behavior—conscious, verbal, intellectual—are functionally supra-summative and demand new specific methods for their approach and study so that both structuralistic introspectionism and objective atomism are necessarily abnegated . . . .
>
> [Furthermore], in dialectical psychology emergence and functional supra-summativeness do not presuppose metaphysical configurational correlates but are held to be merely results of the interaction of parts. The system is thus more similar to the "reaction stages" and "higher units" of Woodworth than to those of Gestalt and of Wheeler and Tolman" (**86**, p. 20).

The above observations appeared in 1935. In 1942 Razran suggested that, all in all, "current Soviet psychology may probably be best designated as a sort of Neo-Functionalism," adding, however, that he thought Soviet psychologists would "no doubt resent such an apellation" (**85**, p. 446). As a matter of fact, they apparently would, for the very reason that "the overcoming of abstract functionalism" (**91**, p. 3) is conceived as one of the major tasks of dialectic psychology. For instance, in 1943 Rubinshteĭn wrote the following:[18]

---

[17] Although in 1924 it was possible to have several papers presented by psychoanalysts before the Second All-Russian Psychoneurological Congress, held in Leningrad, by 1930 the psychoanalytic movement in the U.S.S.R. came to a practical end and the publication of psychoanalytic material, which had already experienced increasing difficulty for a number of years, ceased *in toto*. To this day the hostility to Freudianism and related conceptions continues unremitting. Freudianism is now either ignored or, if noticed, only to be criticized as an "idealistic, metaphysical, and reactionary science, [displaying] the methodological faults of biologism, psychologism, and individualism" (**83**), these three defects characterizing its application to major aspects of man and society.

> Three problems seem to be possessed of an immediacy for psychology at this given stage, and their correct formulation, if not their solution, is one that is especially essential for progressive thinking in psychology.
>
> 1. The problem of the development of the psyche involving, in particular, (a) the overcoming of fatalistic views regarding the development of personality and consciousness; and (b) the problem of development and learning [the problem of development and education].
>
> 2. The problem of actional efficacy and consciousness; in other words, the surmounting of that passive contemplation which reigns in the traditional psychology of consciousness. And in connection with this,
>
> 3. The elimination from psychology of abstract functionalism ["which approaches psychological problems of activity by externally applying abstract principles without consideration of the conditions in which activity takes place"]; and the transition of psychology to the study of the psyche and consciousness in concrete activity, in which the psyche and consciousness are not only made manifest, but also formed.
>
> The shift from the study of functions, only abstractly conceived, to the study of the psyche and consciousness in concrete activity [promises to] enable psychology organically to draw near to the concrete questions of applied practice . . . (**93**).

In view of the foregoing and the basic principles of dialectic psychology, shortly to be enumerated and briefly discussed, it seems to the writer best to regard dialectic psychology as a school of psychology in many respects decisively different from others, instead of hazarding such an ambiguous designation as "neo-functionalism," however useful it may be for suggestive comparisons or hints to assist those of us who are anxious to form some idea of matters where there is so little information available in English. Razran rightly states that "the Soviet psychologist is really not just a laboratory experimentalist, a fact-finder who finds facts to be used somewhere, by some one, some time; [because he] is concerned . . . with the change and control of behavior in a particular direction,[19] [whereas his] American [colleague] is interested in the prediction and control of behavior without emphasis on the direction of the control" (**86**, pp. 19–20). Such a statement in itself is enough to signalize the presence of fundamental differences.

---

[18] In untranslated manuscript of *Bases of General Psychology*, intended for an American edition. The first edition of this important work appeared in 1940 and was awarded in 1942 the Stalin Prize. The writer wishes to acknowledge his indebtedness to Helen Black of the Preslit Literary Agency for the opportunity of studying this manuscript.

A second edition of this work (**96**) appeared in 1946 and was the subject of special discussion in 1947 during a psychological conference. While Kornilov concedes the book to be "monumental" and "first-rate," he reports that among the criticisms leveled at it were (1) insufficient mastery of the dialectical materialist method, (2) conciliatory appeasement of bourgeois psychologists and their theories, and (3) failure to incorporate a study of the psychology of the Soviet citizen (**41**; **47**, p. 13; **117**, p. 31). Leont'ev (**61**), in the course of outling the "tasks" that await Soviet psychology as a result of the recent genetics "debate" from which Lysenko emerged complete victor, also alludes to certain "objectivist" deficiencies inherent in Rubinshteĭn's work.

It is important to realize that not only is dialectic psychology *a* Soviet psychology, but it is *the* Soviet psychology. There is only one school of psychology in the U.S.S.R., and its creation, though admittedly far from complete, is regarded as already an outstanding accomplishment of great significance.

Thus, we find, for example, Rubinshteĭn (**91**) and Kornilov (**46**) referring to "our Soviet psychology" and Anan'ev asserting that "one of the greatest general achievements of Soviet psychology consists in the establishment of theoretical unity within it at the same time that there thrives a tremendous diversity in its creative research . . ." [and that], "with full awareness of the historic role of Soviet psychology, it is necessary to note the fact that Soviet psychology is the first in the whole history of psychology throughout the world to have attained theoretical unity, a theoretical wholeness of psychological theory formulated on the basis of Marxist-Leninist philosophy." [As a matter of fact], "there exists not a single field of psychology, in which there has not been made felt the great transmuting force of the philosophical principles of Marxism-Leninism" (**6**, p. 56). And Teplov, summarizing the achievements of thirty years of Soviet psychology remarks that, "guided by the only true methodology—that of dialectical materialism, Soviet psychology opens an entirely new chapter in the history of psychological science; a chapter in relation to which all previous chapters must appear as only prehistory" (**117**, p. 5).

---

[19] The Soviet psychologist affirms not only the unity, within consciousness, of the social and the personal, but also, among others, the unity of (1) the ideological and the psychological, (2) the moral and the psychological, and (3) the *Weltanschaulich* and the psychological. He asserts, furthermore, that no mischievous "objectivism" must blind the Soviet psychologist to the necessary "unity of psychology and political philosophy" (**9**, **94**).

In a recent article by Rives (**88**), the fundamental interest in man's direction is transparently exhibited in his declarative "study" of not what man is, but what he should and will be in a Soviet society. In other words, in psychology, not man but the "new Soviet man" must be mirrored and projected. It follows, then, that a "complete" psychology of man must go beyond static analyses and provide dynamic models for his emulation.

The Basic Principles

1. *The principle of psychophysical unity.* This principle exhibits the materialist monism of dialectical materialism and involves basically Lenin's theory of reflection. It postulates the unity of the psychic with the internal and the external; the psyche being considered as both a function of an internal organic substratum and a reflector of things in an external world existing prior to and apart from consciousness. Thus, the psychic properties of the brain—"the highest form of organic matter"—are seen to effect the subjective reflection of a real outer world.

This principle, however, is not some pious formula that is to be dutifully invoked and then forgotten as is so often the case elsewhere. Shunning both parallelistic and interactional dualisms, the Soviet psychologist takes the principle, as he sees it, of psychophysical monism very seriously. Thus, for the Soviet psychologist in his investigation of sensations:

> The sensations constitute the object of study, being considered not merely in their functional dependence on the organ, but also in their relation to the object, whose qualities they reflect. In distinction to the purely gnoseological treatment of sensations, which is concerned only with its adequacy or inadequacy to the object, [dialectic] psychology studies sensations as determined by the total life and personality of [an] individual [in action]. Thus, [dialectic] psychological research is concerned not only with the "stimulus" but with the object, too; not only with the organ but with the man (**94**, p. 186).

2. *The principle of the developmental unity of the psyche and the organism*, or *the dialectic principle.* This principle exhibits the famous dialectical triad of thesis, antithesis, and synthesis. Implied are the two unities: the unity of opposites in mutual interconnection and penetration, and the unity of quantity and quality; the latter involving the saltatory *skachók* or emergent "leap" in the process of development (**40**). This principle has, therefore, been also referred to as *the principle of evolution in its dialectical materialist form* (**94**).

The dialectic principle is reciprocal in the sense that it regards the development of the psyche and that of the organism and its mode of existence as reciprocally dependent on each other. Phylogenetically, the development of the psyche is viewed as a by-product in the evolution of organisms, that is to say, as a derivative, though specific, component in this evolution. Thus, in the course of evolution adaptive changes in the form of life and mode of existence are thought to induce modifications not only in the structure but also in the psychophysical functions of the nervous system; but the two of them—structure and function—are conceived always as coexistent in "complex unity, mutual interaction, and interpenetration."

Ontogenetically, the dialectic principle is pivotal in that it affirms the basic proposition that the psychic characteristics and attributes of a personality are both the *premise* and the *result* of behavior and activity as a member *in* and *of* society. This is taken as liberating man from the notion that he is fatalistically predetermined by some all-powerful ineluctable heredity and depersonalized environment that somehow exert immutable effects in fixed sequence; for, according to this principle, the psychic properties of man are in action (education, work, social change, etc.) not only exhibited, but beaten into shape and transformed (62, 63).

Neuropsychic activity is, therefore, held to be a reciprocal function of the organism's form of life. This leads to the statement of two necessary corollaries: (1) The psychic development of animals is determined primarily by biological laws, since animal forms of life are essentially biological; but (2), in contradistinction, man's form of life is not to be equated to that of the animal. It is social, historical, man-created, hence, on a qualitatively different plane; so that the origin and development of human consciousness, itself, is not sought in mere biological genesis, but is instead viewed as being, in the last analysis, determined by the organization of the "means of production," historically evolved and evolving in society.

3. *The principle of historicism.* This principle reflects the historicity fundamental to dialectical and historical materialism. It is applied to the notion of the genesis of human consciousness in the historical process of man's social development. This, too, is a reciprocal conception. Man's social existence and organization, summarizing the long period of his historical evolution, has come to determine his consciousness, his behavior, the form of his thoughts and feelings. Conversely, his consciousness, thoughts, feelings, etc. have acted to condition his social existence and organization.

This implies that the psychology of man must involve into its theory not only his biology, but also his sociology and, further, that he must be regarded concurrently as the end-product of long years of prior historical development and so understood. A psychology that precludes essential interest in his social and historical development is considered as grossly inadequate for man's proper understanding, for the reason that man's social and historical development is looked on as making him subject to psychological laws *qualitatively* different from those characterizing organisms not thus influenced. It is felt that man's psychology, though genetically based on that of his animal predecessors, is not, however, for that reason reducible to it; for the very fact of man's social and historical development is thought to have brought a genuinely new factor into being and established thereby a real discontinuity between the laws governing animal behavior and those governing the specifically human.[20]

---

[20] Thus, in 1939 Anan'ev reported that, while the work of the Department of Psychology in the Bekhterev Institute for Brain Research was defined by the latter's neuropsychological orientation, it aimed, nevertheless, to develop its research problems on the basis of Lenin's theory of reflection and the historical-materialist conception of personality (4).

It is held, further, that even among men, distinctly different laws of the developing psyche must obtain depending on the uniqueness of the social organization which includes these men as active members. Thus, the "class-historical" approach leads to the view that the psychology of the Soviet citizen, for example, and his laws of psychic development are essentially at variance with those of other cultures and that the study of these unique laws must form a special integrated chapter in any psychology that claims to be a complete and truly *Soviet* psychology[21] (**60**, **103**).

4. *The principle of the unity of theory and practice*[22] (stated otherwise, the methodological principle expressing *the unity of research and influence;*[23] or, alternately rendered, *the principle of research and applied action*). This principle asserts the dialectical materialist unity of theory and practice and affirms the Marxian thesis that objectivity of knowledge does not imply passive contemplation on the part of some "neutral" investigator, for the reason that true knowledge about the world is gained not by passive inquiry and speculating about it, but by active intervention and changing it (**109**).

The principle of the unity of theory and practice characterizes, therefore, an experimental methodology which consciously influences, *as it investigates*, the human psyche. This reflects the more general reciprocal thesis that man, in the course of altering the world, thus can get all the more deeply to know it; and, in the course of more deeply knowing it, acquires thereby the greater chance the better to change it.

The human psyche, which is to be investigated under the guidance of this principle, is not contemplated as a static entity bound by psychological laws to exhibit foreordained properties, levels of development, or fixed behavioral patterns because of past biological and environmental factors (**62**, **63**). The human psyche is regarded not as something that can, at the most, merely fulfill limited potentialities set in advance by a constricting inheritance, but as something which can create its own prospects and perspectives for future development through man's own efforts and activities, collective and individual, within his society; heredity and environment being viewed as only factors which condition, but do not dictate, the character, never quite final, of the psyche which an individual may come to display. The words, *the human psyche is in action not only made manifest, but also formed*, is a refrain that with variations is sounded over and over again in many different contexts throughout the whole of not only contemporary Soviet psychology but also other fields of Soviet life (**97**).

---

[21] Concrete manifestation of this view is in abundant evidence in the current glorification of the "new Soviet man" (**88**).

[22] This principle is reminiscent of Dewey's pragmatism and Lewin's "action research," though, no doubt, decided exception would be taken to any assertion of affinity. For example, Smirnov denounces Dewey's ideas as those of a "demogogue" and "most evil enemy of true science and democracy" (**104**, p. 120) and Rubinshteĭn demurs, more moderately, to be sure, against the all-too prevasive "idealism" of Lewin's topological and vector formulations (**62**; **91**, pp. 395, 429).

[23] In a recently translated article (**94**) this unity has been somewhat inadequately referred to as that of "investigation and action (*vozdeĭstvie*)," the parentheticated original indicating that the translator was not quite sure as to how best convey the exact "action" connoted. *Deĭstvie*, minus the prefix *voz*, means generally "action"; *vozdeĭstvie*, however, connotes "action with intent to influence." Moreover, the methodological example, drawn from pedagogy and cited in the same article, makes the kind of "action" intended very clear—"Study the children, teaching them; teach the children, studying them" (**94**, p. 184).

5. *The principle of the unity of consciousness and activity.*[24] This principle reflects the Marxian thesis of the uniqueness of man's consciousness, the inception and development of which is held to be essentially the historical product of labor by man within the social framework. It is, in fact, the grand principle into which the four preceding principles converge.

In this principle consciousness is conceived as formed through concrete activity among men in society, and knowledge of consciousness can be gauged through the activities of people; that is to say, in this very activity consciousness is not only formed, but made accessible to public disclosure and study. Consciousness, therefore, is not the ultimately solipsismal concept that it is for so many, but is an indissoluble unity of the subjective and the objective. In other words, consciousness for the Soviet psychologist, does not exist in a self-contained world locked off unto itself, but exists and is formed in unity and relation with the outer world which it comes to reflect both experientially and cognitionally through man's activity (**95**).

Furthermore, every *experience* of a person, he feels, is always and inescapably an experience of *something*, so that the very internal nature of experience reflects a mediated determination *via* that person's individually grasped or ungrasped relation to the external world of which he is a part. Consciousness is, therefore, looked on as a unity of the direct experientially given and knowledge, the latter not necessarily in any way formalized or verbalized (**91**, p. 57).

Knowledge consists in actively grasped relationships. That which is grasped may be then characterized as partaking of the essence of consciousness; in other words, it may be termed as conscious. Similarly, that which is ungrasped may be termed as unconscious, the latter only in this definite sense. For example, an attitude may exist in an individual and not be *known* to exist by him, that is, he may not grasp (be aware of) his relationship to others and the surrounding world, although it may be very apparent to outside observers. In this contingency we may regard his attitude as one that is unconscious, that is to say, ungrasped by him, though effective in his actions. But an idea, on the other hand, cannot exist in an individual ungrasped, without semantical violence, and, hence, the notion of unconscious ideas existing at odds with conscious man and irrationally exerting power and direction within an individual is held to be untenable (**2, 93**).

---

[24] It is impossible to over-emphasize the role of purposeful action and activity in Soviet dialectic psychology. For example, perception is held to be generated out of activity onto- and phylogenetically. While the stimulus may precede the perception, sensation is a derivative of perception and indicates a high degree of conscious differentiation. In other words, to distinguish sensations is not a primitive but already a highly sophisticated act by a conscious active man (**91**, pp. 151–152).

All this makes also untenable the concept of unconscious wishes, kept essentially *intact*, through the long years in some mysterious subterranean cavern from which to sally forth in one guise or other at appointed occasions to grip and bedevil man; unconscious wishes which subvert and minimize the relevance of man's directed conscious behavior and activity. For the Soviet psychologist, there is no Unconscious, although some aspects of behavior and action are unconscious. In other words, the adjective does not necessitate the substantive, not even for Freudian exigency.

The concept of *consciousness as a unity of the experiential given and knowledge*, gained in action and activity, is one of great moment and consequence in Soviet dialectic psychology. In man the component of knowledge in consciousness is so vastly greater and more organized than could be possibly the case in even his nearest animal relative as to make distinguishable a *qualitative* difference between the psyche of man and that of the next highest on the phylogenetic ladder. Consciousness is, therefore, a peculiarly and specifically human attribute.

Of even greater moment and consequence is the concept of the unity of consciousness and activity, which is, in fact, regarded as the keystone of Soviet psychology. The connection between consciousness and activity is by no means a mere formal device, but is regarded as a deeply intrinsic and reciprocally genetic relationship, whereby the one conditions the advent and development of the other. In other words, through the organization and fashioning of concrete activity it is held possible to influence the formation and essential structure of consciousness, itself; just as consciousness is held to influence the onset and subsequent unfolding of the directed activity that is peculiarly man's. Conscious activity is, therefore, seen as the *sine qua non* of man's psychology.

Accordingly, consciousness and activity are viewed as being in man most intimately and reciprocally related, and the one cannot be understood apart from the other. For example, just as consciousness is related to external behavior, so is external behavior related to consciousness; as the analysis of behavior reveals that the outer aspect of action alone does not define it univocally (**97**). In other words, whereas, on the one hand, the *same* given external movements in one instance may signify in other instances essentially *different* behavior, so, on the other hand, given *dissimilar* external movements at different instances may denote essentially the *same* behavior.

Only by ascertaining the *relation* of man, as a *conscious* being, to his fellow-man and to the world about him may the inner content or meaning of his behavior be gauged. This relationship both constitutes and defines the significance of man's behavior. Stated more generally, the very meaning or content of man's conscious experience and behavior rests on the unity of consciousness and activity.

For these reasons psychology must be understood, according to the Soviet psychologist, as not only *the science of the psyche in developmental action*,[25] but as one whose fundamental unit is the conscious concrete act—not an act along which consciousness meekly tags in psychophysical parallelism, but one that is ontologically different from a non-conscious act or mere mechanical reaction to stimuli.

In this manner the traditional Cartesian dualism of mind vs. body is held to be not only circumvented, but overcome. There is, therefore, no necessity to establish the relationship of behavior, as something external, to consciousness as something that has only internal relevance. Behavior is already in itself, for the Soviet psychologist, a unity of the outer and the inner. Likewise each internal psychological process is a unity of the internal and the external, which in the definitiveness of its content, reflects an internally or, rather, psychologically refracted outer world of sense-infused meaningful objects. Instead of trying to correlate or link an "inactive" consciousness to an "unconscious" activity, the Soviet psychologist unites instead an "active" consciousness with a "conscious" activity.

The basic orientation of the Soviet psychologist, as expressed in the above principles (**9, 93, 94, 119**), force him to go beyond the bounds of a purely functional treatment of the psyche which regards each psychic process as univocally determined from within functional-organic dependencies. They lead him to view each psychic process as mutually dependent on the complex interrelations set up between man and the surrounding world in the course of historically developed social activity. Methodologically, they lead him to the unity of research and influence in his experimental programs. For the Soviet psychologist the psyche is thereby thrust out of its traditional residence in some self-contained inner world. The study of the psyche is, therefore, conducted in the context of the very concrete situations involving the life and activity of real people who, as they manifest their psyche, form and structure it.

---

[25] This final formulation contrasts strongly with those definitions of psychology which made it a science of reflexes (Bekhterev), of behavior (Blonskiĭ and others), or of reactions (Kornilov).

BIBLIOGRAPHY

1. ADRIAN, E. D. Physiology in the U.S.S.R. *Nature*, 1945, **156**, 222.
2. AĬROPETI͡ANT͡S, E., & BYKOV, K. M. Physiological experiments and the psychology of the subconscious. *Phil. phenomenol. Res.*, 1945, **5**, 577–593.
3. ANAN'EV, B. G. Zadachi izuchenii͡a istorii russkoĭ psikhologii. *Sovetsk. Pedag.*, 1938, **2**, No. 4.
4. ANAN'EV, B. G. [Problems of psychological investigation.] *Trud. Inst. Izuch. Mozga Bekht.*, 1939, **9**, 5–14.
5. ANAN'EV, B. G. Iz oboronnogo opyta sovetskoĭ psikhologii. *Sovetsk. Pedag.*, 1943, **7**, No. 2–3, 48–54.
6. ANAN'EV, B. G. Russkai͡a nauchnai͡a psikhologii͡a i eë rol' v mirovoĭ psikhologicheskoĭ nauke. *Sovetsk. Pedag.*, 1945, **9**, No. 3, 47–56.
7. ANAN'EV, B. G. *Ocherki istorii russkoĭ psikhologii XVIII i XIX vekov.* Moscow: Ogiz, 1947.
8. ANAN'EV, B. G. Achievements of Soviet psychologists. *J. gen. Psychol.*, 1948, **38**, 257–262.
9. ANAN'EV, B. G. *Uspekhi sovetskoĭ psikhologii.* Leningrad: Lenizdat, 1948.
10. ANAN'EV, B. G., & SERGEEVA, L. I. Pskhologii͡a pedagogicheskoĭ ot͡senki. *Trud. Inst. Izuch. Mozga Bekht.*, 1935, **4**, 1–142.
11. ARTEMOV, V. A. 25 let sovetskoĭ psikhologii. *Sovetsk. Pedag.*, 1942, **6**, No. 10, 24–30.
12. BABKIN, B. P. Sechenov and Pavlov. *Russ. Rev.*, 1946, **5**, No. 2, 24–35.
13. BABUSHKIN, A. P. Eklektika i reakt͡sionnai͡a kleveta na sovetskogo rebënka i podrostka. *Pedalogii͡a*, 1932, **5**, No. 1–2, 35–41.
14. BABUSHKIN, A. P. Protiv vulgarnogo mekhanit͡sizma v pedologii. *Pedologii͡a*, 1932, **5**, No. 3, 23–30.
15. BASOV, M. I͡A. *Metodika psikhologicheskikh nabli͡udenii nad det'mi.* Moscow: 1926.
16. BEKHTEREV, V. M. *Objektive Psychologie oder Psycho-Reflexologie.* Leipzig: Teubner, 1913.
17. BEKHTEREV, V. M. *Kollektivnai͡a refleksologii͡a.* Petrograd: 1921.
18. BEKHTEREV, V. M. Emotions as somato-mimetic reflexes. In *Feelings and emotions: The Wittenberg Symposium.* Worcester: Clark Univ. Press, 1928. Pp. 270–283.
19. BEKHTEREV, V. M. *General principles of human reflexology.* New York: International Publ., 1933.
20. BLONDEL, C. *Introduction à la psychologie collective.* Paris: Colin, 1928.
21. BLONSKIĬ, P. P. *Ocherk nauchnoĭ psikhologii.* Moscow: GIZ, 1921.
22. BLONSKIĬ, P. P. The subject of psychology and psychopathology from a genetic standpoint. *J. genet. Psychol.*, 1928, **35**, 356–373.
23. BLONSKIĬ, P. P. *Psikhologicheskie ocherki.* Moscow: 1927.
24. BLONSKIĬ, P. P. *Pedologii͡a.* Moscow: GIZ, 1934.
25. BLONSKIĬ, P. P. *Pami͡at i myshlenie.* Moscow: Sotsekgiz, 1935.
26. BLONSKIĬ, P. P. *Razvitie myshlenii͡a shkolnika.* Moscow: Uchpedgiz, 1935.
27. BOROVSKIĬ, V. M. *Psikhologii͡a s tochki zrenii͡a materializma.* Moscow: 1929.
28. BOROVSKIĬ, V. M. Psychology in the U.S.S.R. *J. gen. Psychol.*, 1929, **2**, 177–186.

28a. CHELPANOV, G. I. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1934, **61**, 158.

29. DOBRYNIN, N. F. *Vvedenie v psikhologii͡u.* Moscow: 1929.
30. FEDIN, K. Kino "Akademik Ivan Pavlov." *Pravda*, 1949, No. 53, 3.
31. FOMICHEV, A. Pedagogika. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1939, **44**, 399–450.

32. Fomichev, A. Pedologii͡a. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1939, **44**, 462.
33. Friedl, B. La psychotechnique en URSS. *Monde*, 1936, No. 304.
34. Frolov, Y. P. *Pavlov and his school.* London: Paul, Trench, Trubner, 1937.
35. Galperin, P. I. Zametki o print͡sipe t͡selostnosti. *Sovetsk. Psikhonevrol.*, 1933, **3**, 19–34.
36. Ginet͡sinskiĭ, A. G. Osnovnye napravlenii͡a v fiziologii t͡sentral'noĭ nervnoĭ sistemy. In *Uspekhi biologicheskikh nauk v SSSR za 25 let.* Moscow: USSR Acad. Sci. Publ. House, 1945. Pp. 27–34.
37. Georg'ev, F. Protiv bikheviorizma i reaktologii. *Pod. Znam. Marksizma*, 1937, **1**, 163–169.
38. Jurinetz, W. Psychoanalyse und Marxism. *Unter dem Banner des Marxismus*, Mar. 1925.
39. Kaganov, Z. I. M. Sechenov. In *I. M. Sechenov* (*Izbrannye filosofskie i psikhologicheskie proizvedenii͡a*). Moscow: State Publ. House of Political Lit., 1947. Pp. 3–66.
40. Kedrov, B. M. *O kolichestvenykh i kachestvenykh izmenenii͡akh v prirode.* Moscow: State Publ. House of Political Lit., 1946.
41. Kolbanovskiĭ, V. N. O nekotorykh nedostatkakh knigi prof. S. L. Rubinshteĭna. *Sovetsk. Pedag.*, 1947, **11**, No. 6, 103–110.
42. Kornilov, K. N., *Uchenie o reakt͡sii͡akh cheloveka.* Moscow: 1922.
43. Kornilov, K. N. Sovremennai͡a psikhologii͡a i marksizm. Leningrad: 1924.
44. Kornilov, K. N. *Uchebnik psikhologii, izlozhennoĭ s tochki zrenii͡a dialekticheskogo materializma.* Moscow: 1926.
45. Kornilov, K. N. Psychology in the light of dialectic materialism. In C. Murchison (Ed.), *Psychologies of 1930*, pp. 243–278. Worcester: Clark Univ. Press, 1930.

45a. Kornilov, K. N. *Bol'sh Sovetsk. Ėnt͡sikl.*, 1937, **34**, 324.

46. Kornilov, K. N. *Psikhologii͡a.* Moscow: Uchpedgiz, 1946.
47. Kornilov, K. N. Sostoi͡anie i zadachi nauchno-issledovatel'skoĭ raboty po pedagogicheskim naukam v pedagogicheskikh i uchitel'skikh institutakh. *Sovetsk. Pedag.*, 1947, **11**, No. 9, 8–14.
48. Kornilov, K. N., Smirnov, A. A., & Teplov, B. M. (Eds.) *Psikhologii͡a.* Moscow: Uchpedgiz, 1948.
49. Koshtoi͡ant͡s, Kh. Sechenov. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1945, **51**, 45–52.
50. Kostyleff, N. *La réflexologie et les essais d'une psychologie structurale.* Neuchâtel: Delachaux & Niestlé, 1947.
51. Kravkov, S. V. *Ocherk psikhologii.* Moscow: 1925.
52. Kravkov, S. V. *Glaz i ego rabota.* Moscow: 1932.
53. Kravkov, S. V. Vlii͡anie t͡svetovoĭ adaptatsii glaza na ego reaktivnost' po otnoshenii͡u k inadekvatnym razdrazhitelii͡am. *Dokl. Akad. Nauk. SSSR*, 1949, **64**, 203–206.
54. Lenin, V. I. *Materialism and empirio-criticism.* Vol. 13 of *Collected Works of V. I. Lenin.* New York: International Publishers, 1927.
55. Lenin, V. I. *Aus dem philosophischen Nachlass.* Moscow: 1936.
56. Lenin, V. I. *Filosofskie tetradi.* Moscow: State Publ. House of Political Lit., 1938.
57. Leont'ev, A. N. *Razvitie pami͡ati.* Moscow: 1931.
58. Leont'ev, A. N. Studies in the cultural development of the child. III: The development of voluntary attention in the child. *J. genet. Psychol.*, 1932, **40**, 52–83.
59. Leont'ev, A. N. Emot͡sii. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1933, **64**, 190–

194.

60. LEONT'EV, A. N. Problemy detskoĭ i pedagogicheskoĭ psikhologii. *Sovetsk. Pedag.*, 1947, **11**, No. 6, 103–110.

61. LEONT'EV, A. N. Vazhneĭshie zadachi sovetskoĭ psikhologii v svete itogov sessii Vsesoi͡uznoĭ Akademii sel'skokhozi͡aĭstvennykh nauk im. V. I. Lenina. *Sovetsk. Pedag.*, 1949, **13**, 76–85.

62. LONDON, I. D. The treatment of emotions in contemporary Soviet psychology. *J. gen. Psychol.* (In press.)

63. LONDON, I. D. Theory of emotions in Soviet dialectic psychology. In *Feelings and emotions. Proceedings of the second international symposium. The Mooseheart Symposium.* New York: McGraw-Hill. (In press.)

64. LURII͡A, A. R. The problem of the cultural behavior of the child. *J. genet. Psychol.*, 1928, **35**, 493–506.

65. LURII͡A, A. R. Psychology in Russia. *J. genet. Psychol.*, 1928, **35**, 347–355.

66. LURII͡A, A. R. Die moderne psychologie und der dialektische materialismus. *Unter dem Banner des Marxismus*, Nov. 1928.

67. LURII͡A, A. R. Die moderne russische Psychologie und die Psychoanalyse. *Int. Zsch. f. Psychoanal.*, 1928, **12**, 40–53.

68. LURII͡A, A. R. Die Methode der abbildenden Motorik bei Kommunikation der Systeme und ihre Anwendung auf die Affektpsychologie. *Psychol. Forsch.*, 1929, **12**, 127–179.

69. LURII͡A, A. R. New methods of expressive motor reactions in studying affective traces. *Proc. and Papers 9th Int. Cong. Psychol., New Haven*, 1929, 294–296.

70. LURII͡A, A. R. Die Method der abbildenden Motorik in der Tatbestandsdiagnostik. *Zsch. f. angew. Psychol.*, 1930, **35**, 139–183.

71. LURII͡A, A. R. Krizis burzhuaznoĭ psikhologii. *Psikhologii͡a*, 1932, **5**, No. 1–2, 63–95.

72. LURII͡A, A. R. *The nature of human conflicts.* New York: Liveright, 1932.

73. LURII͡A, A. R. The second psychological expedition to central Asia. *J. genet. Psychol.*, 1934, **44**, 255–259.

74. LURII͡A, A. R. L. S. Vygotskiĭ. *Character and Pers.*, 1935, **3**, 238–240.

75. LURII͡A, A. R. Prof. L. S. Vygotskiĭ. *J. genet. Psychol.*, 1935, **46**, 224–226.

76. LURII͡A, A. R. & LEBEDINSKIĬ, M. S. Die Methode der abbildende Motorik in der Untersuchung der Nervenkranken. *Arch. f. Psychiat.*, 1928, **87**, 471–497.

77. LURII͡A, A. R., & VYGOTSKII, L. S. The function and fate of egocentric speech. *Proc. & Papers 9th Int. Cong. Psychol., New Haven*, 1929, 464–465.

78. MCFARLAND, R. A. Psychological research in Soviet Russia. *Sci. Mon., N. Y.*, 1935, **40**, 177–181.

79. ORBELI, L. A. Akademik Ivan Petrovich Pavlov. *Fiz. Zh.*, 1936, **20**, 199–214.

80. ORBELI, L. A. Puti razvitii͡a fiziologii v SSSR. Uchenie o vyssheĭ nervnoĭ dei͡atel'nosti. In *Uspekhi biologicheskikh nauk v SSSR za 25 let.* Moscow: USSR Acad. Sci. Publ. House, 1945. Pp. 7–27.

81. OSIPOV, V. Bekhterev. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1927, **6**, 90–91.

82. PAVLOV, I. P. *Dvadt͡satiletniĭ opyt ob''ektivnogo izuchenii͡a vyssheĭ nervnoĭ dei͡atel'nosti zhivotnykh.* Petrograd, 1923.

83. PEREPEL, E. The psychoanalytic movement in USSR. *Psychoanalyt. Rev.*, 1939, **26**, 299–300.

84. Pinkevitch, A. *Science and education in the USSR.* London: Gollancz, 1935.

84a. Postanovlenii͡a T͡SK VKP (b), 5 senti͡abri͡a 1931 g.: O nachal'noĭ i sredneĭ shkole. *Spravochnik partiĭnogo rabotnika*, 1934, No. 8, 350–355.

84b. Postanovlenii͡a T͡SK VKP (b), 4 ii͡uli͡a 1936 g.: O pedologicheskikh izvrashchenii͡akh v sisteme Narkomprosov. *Pravda*, 1936, No. 183.

85. Razran, G. H. S. Current psychological theory in the USSR. *Psychol. Bull.*, 1942, **39**, 445–446.

86. Razran, G. H. S. Psychology in the USSR. *J. Philos.*, 1935, **32**, 19–24.

87. Reich, W. Dialektischer Materialismus und Psychoanalyse. *Unter dem Banner des Marxismus*, 1929, No. 7–8.

88. Rives, S. M. Volevye cherti novogo cheloveka. *Sovetsk. Pedag.*, 1948, **12**, No. 1, 12–26.

89. Rubinshteĭn, S. L. Problemy psikhologii v trudakh Karla Marksa. *Sovetsk. Psikhotekh.*, 1934, **7**, No. 1, 1–20.

90. Rubinshteĭn, S. L. *Osnovy psikhologii.* Moscow: 1935.

91. Rubinshteĭn, S. L. *Osnovy obshcheĭ psikhologii.* Moscow: State Pedagogic Publ. House, 1940.

92. Rubinshteĭn, S. L. Psikhologii͡a i pedagogika. *Sovetsk. Pedag.*, 1941, **5**, No. 7–8, 41–47.

93. Rubinshteĭn, S. L. *Osnovy obshcheĭ psikhologii.* (*In manuscripto*, 1943.)

94. Rubinshteĭn, S. L. Soviet psychology in wartime. *Philos. phenom. Res.*, 1944, **5**, 181–198.

95. Rubinshteĭn, S. L. Consciousness in the light of dialectical materialism. *Sci. & Soc.*, 1946, **10**, 252–261.

96. Rubinshteĭn, S. L. *Osnovy obshcheĭ psikhologii.* Moscow: Uchpedgiz, 1946.

97. Rubinshteĭn, S. L. Psychological science and education. *Harv. Ed. Rev.* 1948, **18**, 158–170.

98. Rybnikov, N. A. Istoriografii͡a sovetskoĭ psikhologii. *Sovetsk. Pedag.*, 1943, **7**, No. 1, 39–43.

99. Sakharov, L. [Methods of investigating concepts.] *Psikhologii͡a*, 1930, **3**, No. 3, 3–33.

99a. Samokritika. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1944, **50**, 189–190.

100. Sapir, I. Freudismus, Soziologie, Psychologie. *Unter dem Banner des Marxismus*, Aug. & Dec., 1929; Feb., 1930.

101. Saradzhev, A., Adami͡an, G., *et al.* Dialekticheskiĭ materializm. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1935, **22**, 45–235.

102. Schultz, R. S., & McFarland, R. A. Industrial psychology in the Soviet Union. *J. appl. Psychol.*, 1935, **19**, 265–308.

103. Smirnov, A. A. Zadachi i soderzhanie prepodavanii͡a psikhologii v pedagogicheskikh i uchitel'skikh institutakh. *Sovetsk. Pedag.*, 1947, **11**, No. 12, 60–69.

104. Smirnov, V. Z. Obsuzhdenie knigi prof. E. N. Medynskogo "Istorii͡a pedagogiki." *Sovetsk. Pedag.*, 1948, **12**, No. 5, 118–122.

105. Shnirman, A. L. Bekhterev's reflexological school. In C. Murchison (Ed.), *Psychologies of 1930*, pp. 221–242. Worcester: Clark Univ. Press, 1930.

106. Shnirman, A. L. Present-day tendencies in Russian psychology. *J. gen Psychol.*, 1928, **1**, 397–404.

107. Shnirman, A. L. Second anniversary of the death of Bekhterev. *J. gen. Psychol.*, 1930, **4**, 391.

108. Shore, M. J. *Soviet education; its psychology and philosophy.* New York: Philosophical Library, 1947.

109. Sokolov, M. V., & Fortunatov, G. A. O prepodavanii psikhologii v sredneĭ shkole. *Sovetsk. Pedag.*, 1947, **11**, No. 8, 66–78.

110. Somerville, J. Soviet science and dialectical materialism. *Philos. Sci.*, 1945, **12**, 23–29.

111. Somerville, J. *Soviet philosophy; a study of theory and practice.* New York: Philos. Library, 1946.

112. Spielrein, I. N. Zur Theorie der Psychotechnik. In *Prinzipienfragen der Psychotechnik* by M. Erdelyi, O. Lippman, I. N. Spielrein, & W. Stern. Leipzig: Barth, 1933.

113. Stalin, I. V. *Voprosy leninizma.* Moscow: State Publ. House of Political Lit., 1939.

114. Stalin, I. V. *Dialectical and historical materialism.* New York: International Publishers, 1940.

115. Talankin, A. A. Protiv menshevistvui͡ushchego idealizma v psikhologii. *Psikhologii͡a*, 1932, **5**, No. 1–2, 38–62.

116. Teitelbaum, S. M. The Dalton plan in the Soviet schools. *Harv. ed. Rev.*, 1947, **17**, 91–101.

117. Teplov, B. M. *Sovetskai͡a psikhologicheskai͡a nauka za 30 let.* Moscow: Pravda, 1947.

118. Teplov, B. M. *Psikhologii͡a.* Moscow: State Pedagogic Publ. House, 1948.

119. Tomaszewski, T. O psychologii w Z.S.R.R. *Kwart. Psychol.*, 1947, **13**, 267–315, 333–337.

120. Torbek, V. M. O moikh oshibkakh v knige "Pedologii͡a v doshkolnom vozraste." *Pedologii͡a*, 1932, **5**, No. 1–2, 42–45.

121. Torbek, V. M. Protiv idealisticheskogo biologizotorstva i vulgarnogo mekhanit͡sizma. *Pedologii͡a*, 1932, **5**, No. 3, 31–36.

122. Ukhtomskiĭ, A. A. Dominanta. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1931, **23**, 136–140.

122a. Ukhtomskiĭ, A. A. *Bol'sh. Sovetsk. Ėnit͡skl.*, 1936, **56**, 459.

123. Vavilov, S. I. *Tridt͡sat' let sovetskoĭ nauki.* Moscow: USSR Acad. Sci. Publ. House, 1947.

124. Veĭdemi͡uller, K., & Shchelgov, A. Freĭdizm. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1935, **59**, 189–193.

125. Vinogradov, M. I. Osnovnye napravlenii͡a v razvitii uchenii͡a N.E. Vvedenskogo. In *Uspekhi biologicheskikh nauk v SSSR za 25 let*, p. 39–43. Moscow: USSR Acad. Sci. Publ. House, 1945.

126. Viteles, M. S. Industrial psychology in Russia. *Occup. Psychol.*, 1938, **12**, 85–103.

127. Voĭtonis, N. Bikheviorizm. *Psikhologii͡a*, 1928, **1**, No. 1.

128. Volgin, V. P. Dialekticheskiĭ materializm i istoricheskai͡a nauka. *Izv. Akad. Nauk SSSR, Serii͡a Istorii i Filosofii*, 1945, **2**, 121–128.

129. Vygotskiĭ, L. S. Soznanie kak problema psikhologii povedenii͡a. In *Psikhologii͡a marksizma.* Moscow: 1925.

130. Vygotskiĭ, L. S. Pedagogicheskai͡a psikhologii͡a. Moscow: 1926.

131. Vygotskiĭ, L. S. The problem of the cultural development of the child *J. genet. Psychol.*, 1929, **36**, 415–433.

132. Vygotskiĭ, L. S. *Myshlenie i rech.* Moscow: GIZ, 1934.

133. Vygotskiĭ, L. S. Thought in schizophrenia. *Arch. Neurol. Psychiat.*, 1934, **31**, 163–177.

134. Vygotskiĭ, L. S. Thought and speech. *Psychiatry*, 1939, **2**, 29–54.

135. Watson, J. B. Psychology as the behaviorist views it. *Psychol. Rev.*, 1913, **20**, 158–177.

136. Watson, J. B. Bikheviorizm. *Bol'sh. Sovetsk. Ėnt͡sikl.*, 1927, **6**, 434–443.

137. Zhdanov, A. A. Vystuplenie na diskussii po knige G. F. Aleksandrova "Istorii͡a zapadnoevropeĭskoĭ filosofii." *Bol'shevik*, 1947, **24**, No. 16.

Received December 10, 1948.

STATISTICAL BIAS IN COMPARISONS OF MONAURAL AND BINAURAL THRESHOLDS: BINAURAL SUMMATION OR BINAURAL SUPPLEMENTATION*

MONCRIEFF H. SMITH AND J. C. R. LICKLIDER
Harvard University

In the last decade a number of papers[1] have dealt with the problem of binaural summation at the absolute threshold of hearing: to what extent does central summation of excitation from the two ears make binaural hearing more sensitive than monaural hearing? Since the answer to the question is destined to be ambiguous if the two ears are not equally sensitive, and since few people have matched ears, the procedure has been (1) to measure the two monaural thresholds, (2) to handicap the better ear by an amount equal to the difference between the two monaural measures, and (3) with the ears thus equated to determine the binaural threshold. The difference between the equated monaural thresholds and the binaural threshold, usually about 3 decibels, has been taken as a measure of binaural summation. This procedure was first described by Urbantschitsch (**12**), who used an ear plug to handicap the better ear. In the more recent experiments,[1] attenuators have been used to adjust the currents fed separately to two earphones.

This procedure is analogous, except for the adjustment of the stimuli to the sensitivities of the two receptors, to the one used in comparing monocular with binocular vision; and Pirenne (**9**) and Bárány (**1, 2**) have pointed out that, in the visual case, the procedure by its very nature assures that two eyes will see more clearly than one.

> Let us assume that the visual acuities (or other thresholds) of both eyes fluctuate independently of one another . . . and that the instantaneous thresholds for monocular vision have the same distribution in both eyes, . . . then if the one eye alone has the chance a of seeing the symbol, both eyes together have the chance $2a - a^2$. As a is smaller than 1, this expression will always be greater than a—that is to say, two eyes will be able to see better than one solely as a result of random combination. [Bárány (**2**), p. 127.]

The same source of bias is of course inherent, as Bárány (**1**) has suggested, in the procedure as applied to the determination of thresholds in other sense modalities. It is of interest, therefore, to inquire how great

---

* This report was prepared under contract N5ori-76 between Harvard University and the Office of Naval Research, U. S. Navy (Project Nr 147-201, Report PNR-72). Reproduction for any purpose of the U. S. Government is permitted.

[1] Studies in binaural summation were reviewed in 1948 by I. J. Hirsh (**6**).

the bias is in the monaural-binaural comparison and whether it is influenced by the step of equating the monaural thresholds before determining the binaural threshold.

In order to estimate the magnitude of the bias, it is necessary to define the null condition under which we should say that there is no binaural summation. We can imagine, for this purpose, two monaural listeners, one with only a right ear, the other with only a left ear. The two listeners have no means whatsoever of communicating with each other, but both report to the same experimenter. To obtain measures of monaural and "binaural" sensitivity, the experimenter tests the two listeners separately (successively), then together, in the latter instance recording a positive response whenever *either* listener reports hearing the stimulus tone. It need hardly be said that we do not propose this schema as a psychophysiological hypothesis. There is of course no doubt that the two ears of a single listener do send their messages to a common center in which true summation may occur. Evidence of interaural interaction has been found in psychophysical experiments on sound localization and on masking and in electrophysiological observations at the levels of the cochlear nucleus and the temporal lobe. The schema is therefore a "null hypothesis" which a given set of data should lead us to reject before we say that the data constitute evidence of true binaural summation.

As a first step in estimating the magnitude of the bias (i.e., of the supplementation effect), let us consider a variation of the procedure in which the initial monaural thresholds are determined with very great precision before the two listeners (or ears) are equated and the "binaural" tests made. We can suppose that so many trials are averaged that the initial means are for all intents and purposes the parametric values of the two monaural thresholds. Now, if we assume that in the "binaural" tests the thresholds of the two listeners are independent and are distributed normally and with the same variance about their parametric values, we can reduce the procedure of the monaural-binaural comparison to (a) determining two monaural thresholds, a first and a second, for each listener, (b) finding the difference, again for each listener, between the first and the second (i.e., first minus second), and (c) recording the algebraically larger of the two differences thus found. In as much as the first threshold of each listener was fixed at his parametric mean, the differences were simply the deviations of the second thresholds (with signs reversed) from the parametric means of the distributions of monaural thresholds. It is evident, therefore, that the bias in the monaural-binaural comparison is the bias introduced by selecting the larger of two quantities drawn at random from a normally distributed population, the population of momentary monaural thresholds. It is consequently of interest to examine the sampling distribution of the larger of two random normal variates. This sampling distribution is shown in Fig. 1 in relation to the normal parent distribution.

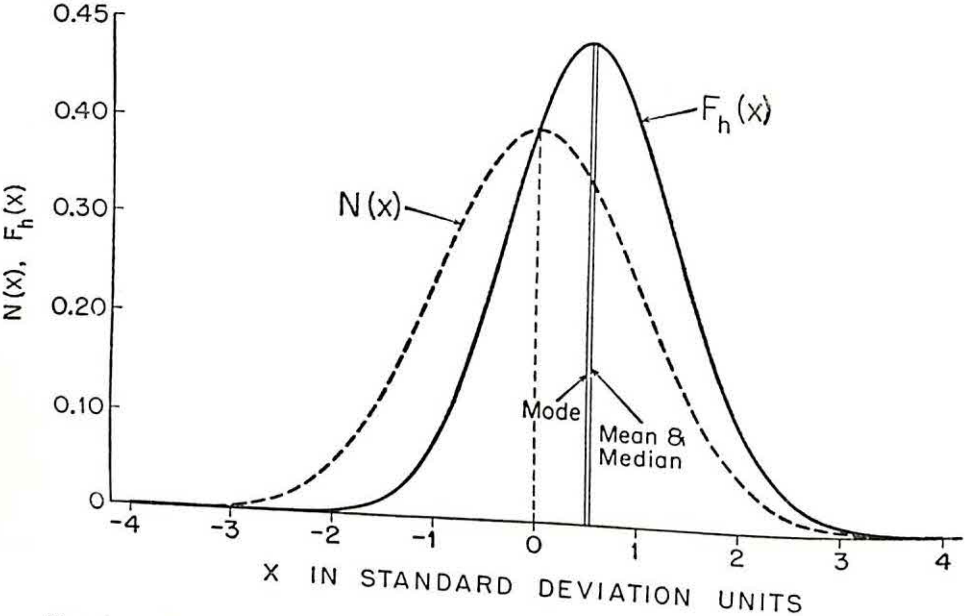


FIG. 1. A NORMAL DISTRIBUTION (DASHED LINE) AND THE SAMPLING DISTRIBUTION (SOLID LINE) OF THE HIGHER OF A PAIR OF VALUES DRAWN FROM IT INDEPENDENTLY AND AT RANDOM.

The distribution-function of the latter is obtained by multiplying the density of the probability that either member of the pair will be x_i by the probability that the other will be less than x_i. If $f_h(x_i)$ is the density of the probability that the first member will be x_i and the second less than x_i, then the sampling distribution-function is

$$F_h(x) = 2f_h(x) = 2\left[\frac{1}{\sqrt{2\pi}}\epsilon^{-x^2/2}\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\epsilon^{-x^2/2}dx\right].$$

Note that $F_h(x)$ is approximately symmetrical. The mean and median lie too close together to be indicated separately in the figure. The mode is approximately 0.06 standard-deviation unit (standard deviation of the normal distribution) to the left of the mean.

The derived distribution is leptokurtic and positively skewed. Hastings, Mosteller, Tukey, and Winsor (5) have determined its standard deviation, which is 0.826 times the standard deviation of the normal parent distribution, and its mean, which is 0.564 standard deviation units higher than the mean of the parent distribution. (The assumption of parent normality appears not to be critical; the corresponding values are 0.817 and 0.577 if the parent distribution is rectangular.) For practical purposes, it is immaterial whether we define our "binaural" threshold as a mean or a median. The mean of the derived distribution is marked by the right-hand edge (0.56) of the line marked "Mean and Median" in the figure, the median (0.55) by the left-hand edge.

An additional component of bias enters the picture as we go a step further and relax the condition that the initial monaural thresholds be determined with great precision. Let us first isolate this new component by assuming that the initial thresholds are averages of n fallible determinations but that the final thresholds are precisely determined in a single momentary trial. Again the bias enters through the recording of the larger of the two differences (first threshold determinations minus second), but the situation is somewhat altered. This time it is the second of the determinations that is assumed to hit the parametric value, and the random variable is the mean of n initial determinations. The amount of bias is therefore 0.56 times the standard deviation of the distribution of means of n determinations. As n increases, the amount of bias decreases. We should not expect the bias to decrease as $1/\sqrt{n}$, however, because the successive determinations would probably not be entirely independent of one another.

The final step is now a simple one. We must relax both restrictions simultaneously and consider both the initial and the final thresholds as final estimates of the parametric value. Once again the bias results from the selection of the larger of two differences, and again it is 0.56 times the standard deviation of the distribution of differences. This time, however, the bias is due in part to supplementation per se and in part to spurious augmentation of the supplementation by error in equating the sensitivities of the two ears. Neither of the terms upon which the difference depends (i.e., neither the initial nor the final threshold) is fixed, and it is more difficult than it was in the restricted cases to express the bias as a function of the standard deviation of the distribution of momentary thresholds.

If we were willing to assume not only independence of the two monaural listeners but also of successive momentary thresholds of a single listener, the bias could be estimated as

$$B = 0.56\sqrt{\sigma_t^2 + \frac{\sigma_t^2}{n}} = 0.56\sigma_t\sqrt{\frac{n+1}{n}}$$

σ_t being the standard deviation of the distribution of monaural thresholds, and n the number of monaural determinations taken in equating the two ears. However, successive momentary thresholds are doubtless not independent. We must have more information about the law of dependence, or we must work directly with the distribution of differences. Even though the distributions of differences for the two listeners may not be exactly equal in variability (the initial thresholds were determined successively, the final thresholds simultaneously, so a longer time elapsed between the two determinations for one listener than for the other), they can be expected to be approximately so, and we shall follow the latter course.

We are now in a position to consider the question: is the bias conceivably great enough to account for the differences between equated monaural thresholds and binaural thresholds that have been found in experiments on binaural "summation"? These differences, for tonal stimuli, have run from 2.5 to 3.6 decibels. Differences obtained with random noise (**10**) have been slightly smaller. Taking 3.0 decibels as a typical value, we see that the standard deviation of the distribution of differences would have to be 3.0/0.56 = 5.4 decibels if binaural supplementation were the only process favoring binaural hearing. How great, then, is the variability of thresholds and of differences between thresholds?

What we should like to know, ideally, is the standard deviation of a distribution of differences between pairs of monaural thresholds, the initial threshold of each pair being based on n determinations of the "momentary" threshold and the final threshold in each pair being based on a single determination. We should hardly expect to find exactly this information in papers on the absolute threshold of hearing, but except for the qualification that the determination should be of "momentary" thresholds, we should not be too far off if we took the standard deviation of a distribution of differences between monaural and binaural thresholds. Hughes (**7**) reports standard deviations (of differences) of approximately 0.5 decibel. Caussé and Chavasse (**3, 4**) give 30 values ranging from 0.78 to 1.22 decibels. Pollack (**10**) shows variances of the sets of differences obtained by individual listeners; the mean of these variances is 2.9, which is the equivalent of a standard deviation of 1.7 decibels. And, finally, Shaw, Newman and Hirsh (**11**) give standard errors of mean differences at eight different frequencies, each value calculated on the assumption that the 26 contributing differences were independent measures; the root mean square of the corresponding standard deviations is 2.4 decibels. If we take these reported standard deviations at their face values, it is clear that 3.0 decibels is considerably greater than the expected value of supplementation. There is in fact room for almost no doubt about the statistical significance of the difference between observed effect and expected bias.

It should be recalled, however, that the expected value of the bias depends upon the standard deviation of the distribution of differences between (1) either individual or averaged values of the initial threshold and (2) *individual* (not averaged) values of the final threshold. We must, therefore, inquire into the natures of the final (binaural) thresholds which entered into the differences, the standard deviations of which we have listed in the preceding paragraph. The psychophysical procedure employed in the experiments on binaural summation has been, without exception, one of the variations of the method of production. The listener has either manipulated the attenuator control himself or has signalled the experimenter which way to adjust it. No time limit has been set for the determination of the values that have been recorded. It would be surprising, therefore, if the quantity recorded were not some kind of an average, a subjective or intuitive average, of momentary thresholds tentatively decided upon during the course of the determination. If such were not the case, it would be difficult to justify the differences in precision among the experiments to which we have referred. It would be difficult to explain how such precision as that indicated by Hughes' 0.5-decibel standard deviation of differences could be obtained when the momentary threshold is found by Lifshitz (8) to vary as much as 6 decibels in 5 seconds.

If each threshold determined by the method of production is, as we believe it to be, more nearly an average than a momentary value, it appears quite probable that the standard deviation of the momentary values that are averaged might be as great as the 5 or 6 decibels required to yield a bias of 3 decibels. We should have to conclude then that the experiments on binaural "summation" at the absolute threshold did not prove the existence of a summative process in binaural hearing.

As matters stand, it is important to have more information on several points before we can reject the hypothesis of binaural supplementation. Three questions are of particular relevance:

1. To what degree are the momentary sensitivities of the two ears correlated? The hypothesis of supplementation has assumed that the correlation is zero, that the two thresholds are independent. If the correlation should prove to be high and positive (bodily noises would tend to influence both thresholds in the same way at the same time; fluctuations of attention might also favor positive correlation), we could expect little supplementation, and it would be easier to demonstrate the existence of a true summative process.

2. To what degree are successive, momentary, monaural thresholds correlated? What we should like to know in this connection is the autocorrelation

function of the monaural threshold. Since the Fourier transform of the autocorrelation function is the power spectrum, we should—with that information—be able to speak precisely about "rates of fluctuation" of the threshold.

3. How does the precision of a threshold determination depend upon the length of time spent in making it? We can regard the listener in a threshold test as a sort of statistical problem-solver, basing an estimate of central tendency (or perhaps of the significance of a difference) on data fed in, in the form of time series, through his sense organs. Adopting this point of view, we see that, if the variability of thresholds determined by the method of production is to have any meaning at all, something has to be said about the length of the time series on which the thresholds are based. A listener with a perfect memory could achieve infinite precision if he were given time enough.

With information on the three questions, it should be possible to decide whether or not experiments on binaural summation yield evidence of binaural summation. Without such information, it can indeed be said that two ears are better than one by about 3 decibels, but the theoretical implications of the experimental results are of considerably less force than they would be if they provided solid ground for rejecting the null hypothesis of supplementation.

BIBLIOGRAPHY

1. Bárány, E. A theory of binocular visual acuity and an analysis of the variability of visual acuity. *Acta ophthal., Kbh.*, 1946, **24**, 63–92.
2. Bárány, E. Some statistical observations on the methods in threshold determinations in general with particular regard to determination of visual acuity and subliminal addition. *Acta ophthal., Kbh.*, 1946, **24**, 113–127.
3. Caussé, R., & Chavasse, P. Récherches sur le seuil de l'audition binauriculaire comparé au seuil monauriculaire. *C. R. Soc. Biol. Paris*, 1941, **85**, 1272–1275.
4. Caussé, R., & Chavasse, P. Différences entre le seuil de l'audition binauriculaire et le seuil monauriculaire en fonction de la fréquence. *C. R. Soc. Biol. Paris*, 1942, **136**, 301–302.
5. Hastings, C., Mosteller, F., Tukey, J. W., & Winsor, C. P. Low moments for small samples; a comparative study of order statistics. *Ann. math. Statist.*, 1947, **18**, 413–426.
6. Hirsh, I. J. Binaural summation—a century of investigation. *Psychol. Bull.*, 1948, **45**, 193–206.
7. Hughes, J. W. The monaural threshold: effect of a subliminal contralateral stimulus. *Proc. roy. Soc.*, 1938, **124B**, 406–420.
8. Lifshitz, S. Fluctuations of the hearing threshold. *J. acoust. Soc. Amer.*, 1939, **11**, 118–121.
9. Pirenne, M. H. Binocular and uniocular threshold of vision. *Nature, Lond.*, 1943, **152**, 698–699.
10. Pollack, I. Monaural and binaural threshold sensitivity for tones and for white noise. *J. acoust. Soc. Amer.*, 1948, **20**, 52–58.
11. Shaw, W. A., Newman, E. B., & Hirsh, I. J. The difference between monaural and binaural thresholds. *J. exp. Psychol.*, 1947, **37**, 229–242.
12. Urbantschitsch, V. Ueber Wechselbeziehungen zwichen beiden Gehörorganen. *Arch. Ohrenhlk.*, 1893, **35**, 1–27.

Received April 4, 1949.

THE GENETICS OF SCHIZOPHRENIA

A SPECIAL REVIEW

NICHOLAS PASTORE
Hunter College

The investigation begun in 1929 and reported by Kallmann[1] in his book, *Genetics of Schizophrenia*, represents the most extensive and ambitious investigation undertaken in recent years in the attempt to demonstrate conclusively a measurable and necessary biological component in the etiology of schizophrenia. This work has been widely cited as having conclusively established a genetic predisposition for schizophrenia, a mental disease which is responsible for the largest proportion of all hospitalized cases for mental illness. For these reasons, a review of Kallmann's methodology is quite desirable. The immediate stimulus for the present review is the recent presentation by Garrison of the conclusions of Kallman's text.[2] This particular presentation contains many errors, both in the quotation and in the interpretation of Kallmann's data, which were quite puzzling to this writer, since Garrison's presentation was intended simply as a summary of Kallmann's major conclusions. It is hoped that the present review will at least give a more accurate picture of the Kallmann investigation.

Kallmann's General Procedure and Conclusions

The object of the Kallmann investigation is to determine the extent and nature of a biological predisposition toward schizophrenia. Kallmann's procedure involves an extended study of 1,087 institutionalized schizophrenics (termed "probands"), their ancestry and descendants. These 1,087 schizophrenics represent the total hospital population for schizophrenia admitted to the Herzberg Hospital, Berlin, in the period 1893 (the year in which the Herzberg Hospital was opened) through 1902. The diagnosis of schizophrenia was made by Kallmann upon examination of the case histories of 15,000 individuals contained in the archives of the hospital (the total number of individuals suffering from all types of mental diseases admitted to the hospital in this ten year period). On the basis of the available data, Kallmann made a diagnosis even though his own diagnosis may have been different from the original diagnosis. This was deemed necessary since "an earlier attempt to sift out the suitable cases of schizophrenia with the help of former records on commitment and diagnosis had been abandoned as useless" (p. 13). Kallmann established his diagnosis "on the basis of neurological and psychiatric findings, the onset and progress of the psychosis and, when possible, on the later findings of autopsy" (p. 13). In those cases "where there was a suspicion of schizophrenia we supplemented the records with whatever material could be obtained from other hospitals and institutions, until we were able to make an unambiguous diagnosis" (p. 13). Furthermore, "all cases in which exogeneous factors might have played a decisive role in the origin or development of a psychosis that resembled schizophrenia" were excluded (p. 14). For example, those cases which seemed to "originate as a clear consequence of tuberculosis" were ruled out.

---

[1] Kallmann, Franz. *The genetics of schizophrenia.* New York: J. J. Augustin, 1938.

[2] Garrison, Jr., Mortimer. The genetics of schizophrenia. *J. abnorm. soc. Psychol.*, 1947, *42*, 122–124.

The ancestry and descendants of the 1,087 selected cases were examined. The total number of individuals, the probands and their ancestors and descendants, involved in this study was 13,851. Since not all of these individuals survived into the danger period for the development of schizophrenia, the actual number involved is perhaps slightly more than one half of the 13,851. A basic part of Kallmann's study involved the calculation of expectancy figures for developing schizophrenia in the children of the probands, in the siblings of the probands, and in the parents of the probands. These expectancy figures for particular family lines were then compared to the presumed expectancy figure of 0.85% (Luxenburger's estimate) for the general population. There was a further subdivision of expectancy figures according to knowledge of the type of mating that was involved. Separate expectancy figures were calculated, for example, for those proband-children who originated in a mating where both parents were schizophrenic, or where one parent was schizophrenic and the other normal, or where one parent was schizophrenic and the other "tainted" by feeblemindedness, psychopathy, and other "not-normal" factors. The subdivision of expectancy figures according to type of mating was intended to disclose additional information with regard to the biological nature of schizophrenia. In line with Kallmann's Mendelian outlook, parents both of whom are schizophrenic should produce only schizophrenic offspring, and a mating in which one parent is schizophrenic and the other normal should produce offspring none of whom is phenotypically schizophrenic. Another major subdivision of his data involved the breakdown of schizophrenia into four clinical forms: hebephrenic (H), catatonic (C), paranoid (P), and simple (S). Separate expectancy figures were calculated for offspring whose schizophrenic parents were classified according to these clinical forms.

A distinctive feature of Kallmann's work is a detailed diagnostic account of the 108 children of probands who were classified as schizophrenic (these proband-children are termed "secondary cases"). This makes it possible for other scientists to evaluate Kallmann's diagnoses of schizophrenic proband-children. It should be mentioned that Kallmann divided these 108 schizophrenics into two categories: definitely schizophrenic (92) and doubtful (16).

Kallmann's major conclusions are as follows:

> 1. For the present we can prove schizophrenia only in a bare 10 per cent of the immediate ancestry of schizophrenic cases; but the total number of ancestors obviously tainted with the hereditary predisposition to schizophrenia and belonging to the heredity-circle of schizophrenia comes, in the direct line, to at least one third of the parents, uncles and aunts (p. 42 f.).
>
> 2. . . . we obtain in the generation of our proband-children the following *total expectancy figures* for the four clinical sub-groups. (p. 107).

	Schizophrenia	Schizoidia
H group:	20.7%	31.9%
C group:	21.6%	30.2%
P group:	10.4%	35.6%
S group:	11.6%	31.7%

> [The term "schizoidia" is used by Kallmann to refer to "eccentric borderline cases" (p. 104).]
>
> 3. . . . the children of hebephrenics and catatonics have about *double* the frequency of schizophrenia of the children of paranoid and simple probands . . . (p. 108). [This difference leads Kallmann to label the hebephrenics and the catatonics as the "nuclear group" and the paranoid and simple schizophrenics as the "peripheral group" (p. 108). Kallmann is inclined to believe that a "biological law" is involved (p. 108).]
>
> 4. . . . the expectation of schizophrenia for an undifferentiated group of proband-children has been found to be 16.4 per cent. . . . The total expectancy figure for siblings of schizophrenics is 11.5 per cent, for half-sisters and half-brothers, 7.6 per cent, for nephews and nieces, 3.9 per cent, and for grandchildren, 4.3 per cent (p. 163).
>
> 5. The probability of schizophrenia in the children of two schizophrenic parents amounts to 68.1 per cent in our survey, while it is only 9.1 per cent for the offspring of the combination of two heterozygotic parents. The cross-breeding proportion of one schizophrenic and one non-schizophrenic parent yields expectancy figures of 14.8 to 23.9 per cent (p. 164).
>
> 6. The genetic connection between both disease groups (schizophrenia and tuberculosis) is proven by the consistent increase of the mortality from tuberculosis in all sub-groups of our survey, and particularly by our finding that this death rate is much higher in the descendants of the nuclear group than in those of the peripheral group, and that it is always highest in the categories with the largest figures for the expectancy of schizophrenia. . . . We . . . discover in all categories of consanguinity such a *startling correspondence between expectation of schizophrenia and mortality from tuberculosis*, that the ratio of frequency be-

> tween the siblings and children of probands, and again between the grandchildren and children, is exactly the same for the expectancy of schizophrenia and for the mortality from tuberculosis.
>
> This statistical result is so unequivocal that it excludes all possibilities of coincidence, and can only be interpreted as a genuine gene-coupling of the tendency to schizophrenia and the heredito-constitutional susceptibility to tuberculous infection. It also indicates an identical pattern of hereditary transmission for these two predispositions, and confirms the assumption that both schizophrenia and tuberculosis represent recessive traits (p. 271).

Although these conclusions are based exclusively on German material, Kallmann thinks that they are also applicable to American population groups (p. xv).

It should not be understood that the foregoing quoted conclusions represent all the major conclusions or that other conclusions are of no interest. The quoted conclusions were selected because they reflect the initial questions underlying the investigation and because they are relevant to the criticisms which will be considered in the next section. Parenthetically, it should be pointed out that Kallmann is deeply interested in eugenic questions and the eugenic implications of his conclusions receive major emphasis in his writing.

KALLMANN'S METHODOLOGY

Establishing the Diagnosis

The diagnoses of the 1,087 probands are based mainly on the clinical records collected by the Herzberg Hospital during the interval 1893–1902 and interpreted by Kallmann more than twenty-five years later. The validity of Kallmann's investigation hinges on the correctness of these diagnoses. First, it should be mentioned, in the way of offering some constructive advice, that Kallmann should have had other psychiatrists check the clinical records (or some sample of records) to determine the consistency of diagnoses. Possibly with this in mind, Kallmann wrote:

> The insistence on an undisputed diagnosis of schizophrenia for all these cases involved exhaustive and time-consuming preparatory labors which were complicated by the fact that case histories of the previous century were naturally recorded under obsolete and erroneous diagnostic headings and were frequently illegible. It was necessary for both the original and secondary cases to be handled and classified by the same person. Then, even if errors in diagnosis were not altogether avoided, they at least would all be along the same line and their effect would be practically uniform, leaving open avenues to correction in the course of the investigation (p. 13).

It could very well be that the diagnostic errors to which Kallmann refers in the preceding quotation consistently emphasize particular diagnostic categories. Furthermore, with only a single scientist rating clinical records, no measure of diagnostic error is possible. This defect could have been remedied, in part, if Kallmann had re-rated a sample of the clinical records after a period of time and in this way determined the reliability of his ratings. A possibly more serious error of indeterminate magnitude involves the reliability of the clinical records themselves. Psychiatrists of the last century must have been guided by the then contemporary diagnostic conceptions and psychiatric theories in observing and recording the behavior of an individual. If this was the case, then the clinical records would unalterably reflect that fact. Kallmann himself seems to have noted this shortcoming when he observed that his "survey" was "limited to case histories that were made decades ago and described according to obsolete medical methods and psychiatric conceptions" (p. 22).

Although an unknown margin of error might have been involved in the classification of the 1,087 probands as schizophrenics (with a possibly additional error by virtue of segregating each of the 1,087 probands into one of the four clinical categories), at least one fact is clear—the probands were institutionalized for mental illness. In classifying the ancestors and descendants of these probands, however, the institutional criterion was completely lacking in a large number of cases. Consequently, many individuals must have been diagnosed on the basis of "anecdotal" materials. This would be especially true for the parents and grandparents of the probands—individuals who probably matured in the first half of the nineteenth century and late eighteenth century. Establishing a particular type of diagnosis for these individuals, almost a century later, seems to be quite risky. From the point of view of genetics, anecdotal information concerning parents and grandparents of probands is worthless, notwithstanding the diligence of the investigator.

Precise knowledge concerning the status of the spouses of the probands would have been of considerable interest, for it would have made it possible to relate the characteristics of the spouses to those of their offspring. Furthermore, even if the status of some of the proband-spouses was in doubt, at least information about their ancestry could have been obtained, as was done in the case of the probands. Kallmann, however, deviated from his previous definitions and classified the proband-spouses in rather unclear terms. For example, one subdivision of proband-spouses places into one category those spouses who are both "schizophrenic or suspicious of schizophrenia" (p. 154). Another single category includes those spouses who are "epileptic, feebleminded, or suffering from an organic nervous disease" (p. 154). It is not known whether institutionalization was an important criterion for the classification of proband-spouses as schizophrenics, as was the case for the probands. Thus, it is unknown whether the term schizophrenia refers to the same set of diagnostic criteria for both proband and proband-spouse.

The diagnosis of schizophrenia in the children of the probands, the "secondary cases," also offers some points for consideration. This group includes 108 secondary cases, of whom 92 were classified as "definite schizophrenic cases" and 16 as "doubtfully schizophrenic cases." The exact status of these "doubtfuls" in Kallmann's work is unknown. Will they all subsequently develop into definite schizophrenics? Should the "doubtfuls" and the "definites" be given equal weight in various calcu- ations? The answer to the first question could not, of course, be conclusive. With regard to the second question, many of Kallmann's major conclusions presuppose that the "doubtfuls" and the "definites" are, in fact, of equal value. Thus, some of his major percentages relating to the expectation of schizophrenia in proband-children actually include the "doubtfuls." For instance, Kallmann's estimate that the expectation of schizophrenia in proband-children is 16.4% subsumes the doubtful cases under the schizophrenic category (see Table 38, p. 112). It is true that sometimes Kallmann carries through two sets of computations, the one including, and the other excluding, these "doubt- ul" cases. In the main, however, he includes these doubtful schizophrenics in the schizophrenic classification. To this writer, there is no ustification for this procedure unless it is assumed that the "doubt- uls" will eventually become definite schizophrenics, in which case the "doubtfuls" should have been so classified at the outset of the investiga- ion.

For the probands, institutionalization for mental disease was a riterion for determining the suitability of individuals for the purposes f Kallmann's investigation. An examination of Kallmann's secondary ases (chapter 5) shows that 35 of the 108 (32%) had not been at all stitutionalized. For an additional few it was not clear whether hos- talization was for a physical or for a mental disease. Furthermore, 6 the 35 noninstitutionalized secondary cases were not alive at the ne that Kallmann began his investigation, thus excluding personal amination of these cases. The fact that the community did not deem necessary to institutionalize a relatively large number of Kallmann's condary cases suggests the possibility that different diagnostic criteria re applied to the probands as compared with the secondary cases.

It would seem at least that different diagnostic criteria were involved in classifying the 6 non-institutionalized secondary cases, who were dead before the investigation began. For these 6 individuals, Kallmann must have utilized the judgments and observations of psychiatrically untrained individuals as a basis for his own classifications.

The methodological purist would have some justification in claiming, with regard to the consistency of diagnostic procedures, that the non-institutionalized cases, and certainly the "doubtfuls," should be excluded from the schizophrenic category. This would reduce Kallmann's 108 secondary cases to 73. A strong case can probably be made for the inclusion of the 35 non-institutionalized individuals but, then, Kallmann's conclusions and inferences should not be set forth with the degree of certainty that Kallmann attaches to them.

Another question remains which deals, not with diagnosis as such, but rather with definitions of the terms "proband" and "secondary case." An examination of the records of the 108 secondary cases shows that 8 of these cases had been institutionalized in the Herzberg Hospital in the interval 1893–1902. One of Kallmann's criteria for selecting probands involved admission into the Herzberg Hospital in the 1893–1902 interval. Consequently, there are 8 secondary cases which, according to the criteria, should be classified also as probands. There are three possibilities: (a) Kallmann did not adhere rigidly to his criteria (a thing which he is entitled to do in this case) and accordingly did not classify the 8 secondary cases as probands. (b) Kallmann classified the 8 secondary cases as probands and excluded their 8 schizophrenic parents from the proband material. If this was done, then Kallmann's figure of 92 definite secondary schizophrenic cases is reduced by 8. (c) Kallmann classified the 8 individuals both as secondary cases and as probands. That is, these 8 individuals were given a double statistical value. If this procedure was followed, then an inflation of incidence figures occurred in two directions. When determining the proportion of proband-children who were schizophrenic, the 8 secondary cases were not classified as probands but simply as secondary cases. When determining the proportion of parents of probands who were schizophrenic, the 8 secondary cases were classified as probands, and then 8 of the probands were shifted into the category of parents.

This last question was considered in detail because Kallmann makes no mention of the point. Doubt is raised as to the exact procedure followed by Kallmann (he should have followed procedure [a] described above) because of his arbitrary statistical treatment of the following analogous issue. Three secondary cases had parents both of whom were probands. Kallmann gave these 3 secondary cases a "double statistical value," that is, they were "counted twice," in calculating the expectation of schizophrenia in proband-children (p. 165). This dubious procedure, of course, increases the size of the expectation and incidence figures. (In our discussions up to this point, we have been citing the figure of 108 secondary cases. Actually, Kallmann speaks of 111 secondary cases—the 108 secondary cases plus the 3 cases of double statistical value.)

Sampling Difficulties

For a problem such as the one at hand, it is very difficult, if not impossible, to obtain a random sample. One can receive some assurance that a sample is not unduly affected by selective factors by comparing some characteristics (characteristics which are independent of schizophrenia) of the sample with those of the general population. The characteristics chosen for comparison are themselves selected, knowingly or otherwise, from a larger population of possible characteristics. Therefore, the coincidence of some selected sample characteristic with the population value should not lead one to conclude that the sample is random or that the sample is not affected by selective factors. On this point, Kallmann claims that his sample was "random" (p. 7) when, in fact, it was not. By the word random he probably meant that the values of certain sample characteristics (for example, male-female ratio) coincided more or less with those of the population at large. In the interests of clarity, Kallmann should not use an already precisely defined term, that is, "random," in a different sense. Moreover, it should be mentioned that for the particular characteristics discussed by Kallmann, the values prevailing in the general schizophrenic population are actually unknown. In choosing the sample characteristics, Kallmann should have considered other properties relevant to his problem. For instance, the question of whether the distribution of his schizophrenic individuals according to the form of schizophrenia was the same as the one which prevailed for the general population of schizophrenics should have received some discussion.

The foregoing general methodological considerations should have served, at least, as a note of caution in generalizing his conclusions from the available data. Kallmann, however, asserts his major conclusions as unassailable findings. The presence of possible selective factors governing the inclusion of individuals in his sample, should have served as a further note of caution. These possible selective factors will receive some brief discussion.

Proband material. We cannot neglect the possibility that, generally, only the more severe cases were institutionalized (and, in particular, at the Herzberg Hospital), and that such cases were more likely to have an inherited component.

There were marked age differences among the probands. The range of the birth-dates extended from 1820 to 1887 (p. 24). These differences were reflected in the wide age differences which prevailed among the probands classified according to the form of schizophrenia (cf. table 3, p. 24).

A special selective factor, which affects the inclusion of all probands, concerns the relative proportion of remissions among the various schizophrenic disease-forms. For example, those aged schizophrenics who were institutionalized for decades probably had the smallest proportion of remissions and this, in turn, might be related to the relative incidence of schizophrenia in their offspring.

The fact that Kallmann "ruled out all cases in which exogenous factors might have played a decisive role in the origin or development of a psychosis that resembled schizophrenia" (p. 14) seems to be a contributory selective factor. For example, "wherever we found tuberculosis we made the diagnosis of schizophrenia dependent on the question of whether the psychotic symptoms occurred before or after the active tuberculosis had broken out. We tried, at least, to establish that schizophrenia did not originate as a clear consequence of the tuberculosis" (p. 14). Thus, Kallmann excludes those schizophrenias which, according to his statement, are environmental in origin. If such schizophrenics had been included in his calculations, the effect probably would have been to decrease the expectation of schizophrenia in the offspring.

After having chosen 1,087 probands, Kallmann "dropped 40 cases" because "some new doubts arose . . . as to the unassailable certainty of the original diagnosis of schizophrenia" (p. 15). Kallmann provides no definite information on the nature of these doubts. Furthermore, although these 40 cases "produced a series of undeniable secondary cases of schizophrenia" (p. 15), it would be of interest to compare the incidence of schizophrenia in the offspring of these 40 probands with that of the offspring of the other probands. Kallmann sometimes includes the 40 cases in his calculations and summaries without justifying their inclusion.

Proband-children. It should be mentioned that such children were the offspring of a selected number of probands, namely, those probands who were fertile. Inasmuch as 489 probands of the original 1,047 probands produced children[3] (there were 510 married probands) it would be of interest to know whether married and unmarried probands differed with respect to certain relevant characteristics. For instance, is the ancestry of the unmarried probands more "tainted" than the ancestry of the married probands, or, is the distribution of "tainted" siblings proportionately the same for both groups?

The married probands produced 2,000 children, of whom approximately 1,000 died before having reached the twentieth year. Although it is not possible to get any information on such deceased individuals, the selection of data which is thereby introduced should have served as another reason for caution in establishing rigorous conclusions.

The sample consisting of 111 secondary cases is too small for the intensive statistical analysis which is involved (a specific example of this point is discussed in the next section under [1]).

---

[3] See table 12, p. 52 for these figures.

Proband-siblings, proband-parents, proband-grandparents. It suffices to point out that the various expectancy figures involving these categories are based on incomplete data; the data were probably least complete for proband-grandparents. The effect of such selection of individuals (individuals were selected because of availability of information concerning them) upon the various expectancy figures is quite unknown.

Tabular and Statistical Presentation of Data

For a ready and complete understanding of his data, the Kallmann monograph is generally inadequate: the tabular and statistical presentation of the data involve some shortcomings.

1. Results are sometimes given in terms of percentages without revealing the sizes of the sub-groups on which the percentages are based: in some cases, numbers are too few to warrant any generalizations (cf. p. 112, table 38). Table 38 deals with the "differentiated expectation of schizophrenia and schizoidia in all proband-children." In this table, the expectancy figures for schizophrenia fall into 24 different categories—that is, there are 24 cells among which the schizophrenic proband-children are distributed for purposes of calculating the various expectation figures. Since there are 111 schizophrenic children all told, it follows that, on the average, each of the 24 expectancy figures is based on 4½ schizophrenic children. The breakdown involving children of simple-schizophrenic probands is not at all meaningful since these probands produced only 4 schizophrenic children.

2. Overlapping categories are sometimes used. For example, in table 10, p. 40, which deals with the "incidence of schizophrenia" in the ancestry of legitimate probands, the following two categories overlap: "schizophrenia in one parent," and "schizophrenia in one parent and other abnormalities in the second parent."

3. The number of schizophrenic proband-children is not 111. The figure of 111 covers 3 individuals who were counted twice, 8 individuals who were possibly classified as probands, 16 individuals classified as "doubtfuls," 35 non-institutionalized individuals[4] (see pp. 291 ff. *supra* for further discussion of these points). One definite schizophrenic secondary case is classified as an "incipient schizophrenic" (p. 191, case #89).

4. Kallmann consistently chooses the higher expectation and incidence figures, where a choice is possible. One factor which is partly responsible for this differential choice is the method of dealing with the "definite" and "doubtful" schizophrenics. As stated before, Kallmann sometimes carries through two sets of computations, one including, the other excluding, the "doubtfuls." In reporting his over-all conclusions, however, the figures which include the "doubtfuls" are presented as the final figures. For example, with reference to table 48, p. 129, the expectation of schizophrenia in proband-siblings is 7.5%, if the

---

[4] The 35 non-institutionalized individuals includes the 16 "doubtfuls." One individual, due to a typographical error, obviously has a wrong birth-date in the record (p. 180, case #56).

"doubtfuls" are excluded, and 11.5% if the "doubtfuls" are included. In presenting his overall conclusions, the 11.5% figure is cited as the expectation of schizophrenia in proband-siblings (see p. 287 *supra* for a quotation of Kallmann's conclusion on this point). A similar situation prevails in the calculations of the expectation figures for proband-children (see pp. 106 ff.).

Another factor which is partly responsible for Kallmann's differential choices may be the fact that there exists no generally accepted objective method for calculating expectancy figures. In chapter 4, Kallmann calculates the expectation of schizophrenia in proband-children in three ways—two of the methods were developed by Weinberg and the other by Ilse. The two Weinberg methods yield expectations of 14.4% and 16.4% and the Ilse method yields an expectation of 12.8% (pp. 139–141). The Ilse method, according to Kallmann, has certain objectionable features which make it unsuitable for the purposes at hand. Of the two Weinberg methods, the method which Kallmann recognizes as "doubtless the more accurate one" (p. 139) yields the expectation figure of 14.4%. The other Weinberg method (the abridged method) yields a figure of 16.4% which is the figure cited by Kallmann in his over-all conclusions, and throughout the book. Kallmann's reason for choosing the 16.4% figure is that the method for its calculation is less complex than the other Weinberg method and the fact that "the final results in both methods correspond so closely that they may be taken as sufficient evidence of the statistical reliability of all our other expectancy figures calculated by the abridged method" (p. 139). In the interest of accuracy, Kallmann's procedure seems quite dubious—if the method which yields a 14.4% figure is the more accurate one then that is the method which should be used regardless of its complexity. Moreover, the closeness of two results (the 14.4% and 16.4% figures) for calculating a single expectancy figure is not a sufficient guarantee that such closeness will obtain in other cases. It was the disparity between the two Weinberg methods which led Kallmann to use the Ilse method which, in turn, was rejected (pp. 139 ff.).

5. At least one basic table cannot be understood by the writer as it stands. Table 58, p. 157, deals with the "expectation of schizophrenia and schizoidia in the offspring of two schizophrenics." The expectation that such offspring will develop schizophrenia is 68.1% and the expectation that they will develop schizoidia is 45.7%. It would seem, then, that the probability that the offspring of parents both of whom are schizophrenic will be either schizophrenic or schizoidic exceeds 100%. It is possible, of course, that Kallmann has overlapping categories in this particular case. But Kallmann does not discuss this point and there seems to be little justification for overlapping categories in this case. If schizophrenic offspring were initially schizoidic then they should be classified only under the schizophrenic category.

6. Occasionally the numbers which appear within the body of a table are not consistent with other statements (a conspicuous example of this kind is discussed below). On page 12, in discussing the selection of individuals as probands, Kallmann states as a criterion of selection: "They must have been schizophrenics beyond the possibility of doubt, and the disease must have become manifest before the age of forty." Yet in table 53, page 140, which gives the "age distribution of probands at disease-onset," it is seen that disease-onset for 27 probands was at the age of 40 years or later. Furthermore, 3 schizophrenic proband-children had a disease onset in the fortieth year or beyond.

7. Legends at head of tables are sometimes inaccurate (see the following discussion).

INCIDENCE OF SCHIZOPHRENIA IN THE ANCESTRY OF LEGITIMATE PROBANDS
(Adapted from Kallman's Table 10, p. 40.)

	Number of Legitimate Probands		*Definite Schizophrenia in a Parent, Grandparent, Aunt or Uncle (percentages)*					
			(*1*)	(*2*)	(*3*)	(*4*)	(*5*)	
Types of Schizophrenia	*Absolute No.*	*Per cent*	*Schizophrenia in one parent*	*Schizophrenia in both parents*	*Schizophrenia in one parent and other abnormalities in the second parent*	*Schizophrenia in a grandparent*	*Schizophrenia in a sibling of the parents*	*Total*
H	414	84	3.9	0.2	2.2	1.0	2.9	10.1
P	225	89	3.1	0.9	1.8	2.7	2.7	11.1
C	136	90.7	2.2	—	0.7	1.5	2.9	7.4
S	141	93.4	1.4	—	1.4	1.4	1.4	5.7
Total No.	916	87.5	3.1	0.3	1.7	1.5	2.6	9.3

In interpreting Table 10, Kallmann writes, "Of the *legitimate* probands, roughly one third have a schizophrenic or schizoid parent, grandparent, aunt or uncle" (p. 38). Further, he writes, "On the average, not even a tenth of our probands (only 9.3 per cent) had a parent or another direct ancestor who was a proved schizophrenic" (p. 39). Both of these statements are apt to be misconstrued. They *do not* refer to the percentage of the ancestors of the probands who are either schizophrenic or schizoid. Rather, they are intended to indicate the percentage of the probands who have either a schizophrenic or a schizoid ancestor. In other words, the base of the various percentage figures is not the number of ancestors but rather the number of legitimate probands. Although Kallmann is probably aware of this fact, yet he slips into the error of presenting various percentage figures with their incorrect base.

It should be noted that the legend of Table 10 does not correspond to the body of the table. The legend refers to the "incidence of schizophrenia in the ancestry of legitimate probands," whereas the last row of numbers refers to the percentage of probands who have schizophrenic ancestors. Kallmann falls into the same error when he concludes that "no more than 10.4 per cent of the proband-parents could be ascertained to have been definitely schizophrenic" (p. 164). This error just doubles the number of schizophrenic parents of the probands. Of the 916 legitimate probands, 3.1% of them had one schizophrenic parent and 0.3% of them had two schizophrenic parents. The total number of schizophrenic parents is therefore, at most, 31 (3.4% of 916). On the other hand, if the number of parents (916×2) is incorrectly used as the base for the 3.4% figure, then the resulting figure would be 62. Similarly, this error would quadruple the number of schizophrenic grandparents of the probands.

Another error exists in this table. Kallmann has overlapping categories which have the effect of raising the percentages. (Compare the legends in columns 1 and 2 with the one in column 3.)

The 9.3% figure has a valid interpretation (ignoring the question of overlapping categories), viz., 9.3% of the probands, roughly 90 probands, will turn up one schizophrenic ancestor, if his ancestry is examined in the manner prescribed by Kallmann. But, then, if the ancestry of the probands is examined sufficiently carefully and extensively, all probands will almost certainly lead to a schizophrenic ancestor.

Consider now Kallmann's conclusion that 10.4%[5] of the proband-parents were definitely schizophrenic (this conclusion was not included in the list cited on pp. 287–288.) For one thing, the term parent is used to cover ancestors in general. Also, "other abnormalities" (see column 3) are classified under schizophrenia. In any case, the correct percentage to convey the idea which Kallmann intends is 1.7% (31 schizophrenic parents distributed among 1,832 parents). This percentage is considerably reduced if all the investigated ancestors of the probands are listed under the term of 'parents,' as is done by Kallmann.

This section refers to Tables 34–37, pp. 106–110, which deal with the "probability of schizophrenia and schizoidia" in the children of the probands. Separate probability figures are also given for proband-children depending on the form of schizophrenia which characterizes the probands. According to these tables, the probands produced a total of 111 children who were definitely or doubtfully schizophrenic. These 111 children are distributed according to the form of schizophrenia of the proband. In chapter 5 of the book, Kallmann presents an account of each proband-child, making it possible to arrive at another estimate of the distribution of the 111 children. The two distributions, the one based on the data presented in Tables 34–37 and the other based on chapter 5, are compared in the following table.

---

[5] It is not stated how the figure of 10.4% was obtained. It probably represents, however, a weighted average of two percentages, the one based on legitimate probands, the other on illegitimate probands.

Form of Schizophrenia in the Proband	*Distribution of the Number of Schizophrenic Children According to Form of Schizophrenia in Proband*	
	(Tables 34–37)	*(Chapter 5)*
Hebephrenia	55	55
Catatonia	23	23
Paranoia	22	29
Simple	11	4
Total No.	111	111

The discrepancies between the two distributions constitute a serious weakness in Kallmann's conclusions since many percentages are based upon figures given in the column entitled "Tables 34–37" in the above table. There is reason to believe that the distribution obtained from chapter 5 is the accurate one, for each individual schizophrenic proband-child is described and classified according to the type of schizophrenia in the proband. The discrepancy is especially significant in so far as the S group is concerned. Kallmann's expectancy figure for the proband-children whose parents fall in this group is 11.6% (p. 107). This percentage is based on 11 schizophrenic children, whereas the actual figure should be 4. With a correction for this error, the 11.6% is reduced to about 4.0%. Parenthetically, female simple-schizophrenic probands produced no children of like mental disease or any mental disease (according to chapter 5). Table 37, p. 110, indicates 8 simple schizophrenic children were produced by such probands.

This paragraph deals with Tables 44–47 (pp. 125–128) which give the "probability of schizophrenia and schizoidia" in the siblings of the probands. The number of siblings involved in the various calculations is close to 2,000. The tables provide information as to the number of siblings who originate in five different types of matings. The two matings of relevance are those which involve "one parent schizophrenic, the other normal or psychopathic" and "one parent schizophrenic, the other schizoid or suspicious of schizophrenia" (p. 125). These two types of matings produced 366 of the 2,000 siblings. On the basis of these data, it is possible to form a minimum estimate of the total number of parents of the probands who were schizophrenic. The average number of children per family produced by the parents is 4.5, the division of 3.5 into 366 yields about 104 family units wherein each family unit contains a schizophrenic parent. (In the preceding calculation, 3.5 was used as the divisor since one of the offspring in a family unit is a proband.) This estimate of 104 schizophrenic parents contradicts the value of 31 schizophrenic parents which was obtained from Table 10, p. 40. This contradiction calls for a clarification.

EVALUATION OF KALLMANN'S MAJOR CONCLUSIONS

Kallmann's major conclusions were presented in an earlier section of this review. Each of these will now be considered in light of the foregoing discussion of Kallmann's methodology. The numbers in the following discussion correspond to the numbering of the conclusions on page 287.[6]

1. "*For the present we can prove schizophrenia only in a bare 10 per cent of the immediate ancestry of schizophrenic cases.*" On the basis of his own data, this conclusion is incorrect. The ten per cent figure does not tell us anything about the proportion of proband-ancestors who were schizophrenic. For this percentage is not based on the total number of ancestors, but rather, on the number of probands who have a schizophrenic parent or ancestor in the family. The actual incidence figure for schizophrenia in proband-ancestors is much less than 2%. Similar comments apply to the second part of conclusion (1). In any case, it is questionable whether the data pertaining to proband-ancestors is sufficiently reliable to warrant the calculation of incidence figures. The diagnosis of proband-parents and proband-grandparents, of individuals who lived in the last half of the eighteenth century and the first half of the nineteenth century, is, at best, tenuous.

2. *The total expectancy figure for schizophrenia in the children of the simple schizophrenic probands is 11.6 per cent.* An actual count of the number of schizophrenic proband children falling into this particular parental sub-group yields an incidence figure of approximately 4 per cent.

3. *The probands can be divided into two distinct schizophrenic groups—"nuclear" and "peripheral."* With the correction for conclusion (2) in mind, this distinction between "nuclear" and "peripheral" schizophrenia becomes confused.

4. "*The expectation of schizophrenia for an undifferentiated group of proband-children has been found to be 16.4 per cent.*" The schizophrenic category, here, includes the "doubtfuls." Therefore the 16.4 percentage is an inflated figure. Furthermore, the 16.4% figure is based on a method which admittedly is not as accurate as another method which yields a figure of 14.4%. The diagnostic criteria for schizophrenia in the various groups mentioned probably do not relate to the same set of facts.

5. "*The probability of schizophrenia in the children of two schizophrenic parents amounts to 68.1 per cent in our survey, while it is only 9.1 per cent for the offspring of the combination of two heterozygotic parents.*" The 68.1% figure is calculated

---

[6] For continuity of presentation, these conclusions are either briefly quoted or restated in this discussion.

from data which yield an expectancy figure significantly larger than 100% that offspring of schizophrenic parentage will be either schizophrenic or schizoid. It is possible that overlapping categories, for which there is little justification, are involved. The data on which the 9.1% figure is based must refer to proband-parents and/or proband-siblings and their spouses—individuals for whom diagnostic errors, and errors due to partial accumulation of data, are the largest.

6. "*The ratio of frequency between the siblings and children of probands, and again between the grandchildren and children, is exactly the same for the expectancy of schizophrenia and for the mortality from tuberculosis.*" *This result indicates a "genuine gene-coupling of the tendency to schizophrenia and the heredito-constitutional susceptibility to tuberculous infection.*" The constancy of the two sets of ratios involves several questionable procedures. First of all, Kallmann's categories for schizophrenia in these calculations subsume the "doubtful" cases. The omission of such cases would completely upset Kallmann's conclusions (see Table 76, p. 246). Second of all, substitution of the more accurate figure of 14.4% for the 16.4% figure in calculating the ratios would also lead to conclusions contradictory to those stated by Kallmann. These two facts alone indicate that his results are not "unequivocal." Assuming the correctness of Kallmann's ratios, the reasoning which leads to the conclusion of a "genuine gene-coupling" between the two diseases is not set forth. That this reasoning is probably invalid, whatever it happens to be, is indicated by the fact that the calculations for the expectation of schizophrenia and mortality from tuberculosis do not involve the same individuals.

A striking result in Kallmann's investigation (not brought out by Kallmann) is the finding that the offspring of probands developed only schizophrenia and no other psychosis. Furthermore, the schizophrenic offspring of probands develop, on the average, the same form of schizophrenia as their parents. Thus, an examination of the records of the secondary cases in chapter 5 shows that 51 of the 92 definite schizophrenic secondary cases (54%) develop the same form of schizophrenia as that which characterizes the parent. There are wide differences in this correspondence according to the disease form. Thus, hebephrenic female probands produced a total of 32 definitely schizophrenic offspring of whom 26 (81%) had the corresponding disease-form. The corresponding figure for schizophrenic offspring of hebephrenic male probands is 6 out of 14 (43%).[7] Kallmann's data indicate that for each disease form there is a closer correspondence between female probands and offspring than between male probands and their offspring.

The above evidence would suggest, in line with Kallmann's methodology, that there is a specific hereditary predisposition for each disease form, and that there are sex differences for such specific hereditary predispositions.

---

[7] Although from the point of view of the present writer the number of cases is too small to establish fully reliable differences among these various groups, this discussion is included because Kallmann frequently assumes the reliability of differences based on the same, or even smaller, number of cases. The writer simply wishes to point out that certain relevant breakdowns are possible, from the standpoint of Kallmann's methodology.

Kallmann's general scientific attitude with regard to the relative weight of genetic and environmental factors in the etiology of schizophrenia is the conception that a demonstrated necessary hereditary factor does not preclude an environmental component. Kallmann's methodological procedure, however, implies that environmental differences are not related to the etiology of schizophrenia. Kallmann interprets various expectancy figures[8] as supporting a predictable Mendelian pattern. Since environment is certainly a variable in Kallmann's data, and is not necessarily correlated with genetic predisposition, a Mendelian methodology in this case, implies that environmental differences are without effect on the development of schizophrenia.

A specific environmental variable uncontrolled in Kallmann's investigation is intactness of family structure. This variable can be related to certain differences in expectancy figures which Kallmann obtained. In this connection, it should be recalled that Kallmann distinguishes between two types of schizophrenias—"nuclear" group which includes hebephrenics and catatonics and the "peripheral" group which includes the paranoids and simple-schizophrenics. The basis for this distinction was the finding that the expectancy figure for the offspring of the "nuclear" probands is double the expectancy figure for the offspring of the "peripheral" probands (p. 108). Kallmann interprets the difference between the two groups along biological lines. This difference, however, can be interpreted in terms of differing degrees of intactness of family structure. The "nuclear" group is characterized by a longer period of hospitalization and an earlier disease onset than the "peripheral" group (see Table 4, p. 26). Both of these factors would contribute to a greater degree of impairment in the family structure of the "nuclear" probands and their offspring. In a different connection, Kallmann recognizes the fact that the "family structure" of paranoid probands is "generally not severely affected and rarely breaks down altogether when the psychosis in one of the parents occurs comparatively late" (p. 10). It was this statement, in fact, which suggested to the present writer the possibility that intactness of family structure is a relevant and uncontrolled variable in Kallmann's investigation. It should be mentioned, furthermore, that intactness of family structure

[8] Over-all expectancy figures, expectancy figures for offspring of parents of a particular form of schizophrenia, expectancy figures as related to the presumed heterozygosity or homozygosity of parents.

parallels the different types of biological matings. Thus, the family structures in which both parents are schizophrenics will show the largest amount of impairment.

SUMMARY

In this review, the writer has dealt only with those points which to him seemed to be of major importance. Thus, discussion of the relationship of Kallmann's eugenic suggestions to his obtained results was omitted despite the fact that eugenic considerations were foremost in Kallmann's thinking. After consideration of the various questions which can be raised with regard to Kallmann's methodology—his diagnostic procedures, statistical treatment of data, sampling procedure, uncontrolled variables—the present writer is of the opinion that the genetics of schizophrenia is still an open question. It should be emphasized that the present writer in no way wishes to minimize the contribution of hereditary factors in the etiology of schizophrenia, however unknown the specific nature of these hereditary factors may be. The Kallmann investigation, however, supplies no reliable information for assessing the genetic basis of schizophrenia.

Received February 12, 1949.

NOTES ON "THE VALIDITY OF PERSONALITY INVENTORIES IN MILITARY PRACTICE," BY ELLIS AND CONRAD*

DONCASTER G. HUMM AND KATHRYN A. HUMM
Los Angeles, California

The importance of the problem of measuring personality is evidenced both by the attempts which are being made to solve it and by the published discussions of these attempts.

The currently available standardized measures of personality have disappointed a number of psychologists who have attempted to use them, and the attempts of these psychologists and the discussions of their findings in the literature have been equally disappointing to many working in this field.

The review by Ellis and Conrad of a number of reports of the military use of personality questionnaires seems to the writers to perform a service in calling attention to a number of faults in making assumptions, in controlling criteria, and in applying mathematics; but this review includes some statements which seem to us to indicate failures to appreciate the nature of personality questionnaires—failures as serious as those of the military users.

This is clearly shown by their conclusions, which we feel need discussion.

Their first conclusion is that temperament measures will be effective only if they are "specially designed for the group to whom they are applied." The standardization of such an instrument, involving as it does a case study of every subject to be used, the tabulation of the answers of all standardization subjects classified by the temperamental groups into which they fall, the analysis of the answers to each question to determine which temperamental groups it distinguishes, the further analysis of the results to establish the effect of response-bias and to compute compensations for it, the validation of the completed questionnaire by trying it on a new group of subjects again classified by case study according to temperamental groups, and finally the administration of the questionnaire to a representative population to determine norms—all this is a laborious and expensive process.

The repetition of such labor and expense for each group to be tested would be justifiable, if feasible at all, if it would actually improve the results; but, on the contrary, the more one instrument can be used in

---

* Ellis, A., and Conrad, H. S. The validity of personality inventories in military practice. *Psychol. Bull.*, 1948, **45**, 385–426.

ifferent situations, the more data can be compared and the more can
e learned about the likenesses and differences among people.

Our experience leads us to the firm conviction that there are person-
lity characteristics which can be identified by the use of a well-stand-
rdized questionnaire; that they are to be found in various combina-
ions throughout the population; that their manifestations may be de-
irable or undesirable according to the combinations in which they occur
nd according to the situations to which their possessors are exposed;
nd that the general application of an instrument which measures them
o a wide variety of groups is an important way of advancing our knowl-
dge of personality.

The reviewers' second conclusion, that "special attention should be
iven to persuading or inducing respondents to answer the inventory
ems as truthfully as they can," is simply naive. It would, of course,
reatly simplify the measurement of personality if all subjects could
e induced to respond with equal "honesty." This cannot be expected,
owever, in many situations in which such instruments are most needed.
Ioreover, the problem is not simply one of "honesty"—some subjects
eem to be "too honest" and to claim disabilities they do not have.
Vhile the response-bias of the same subjects may differ from one testing
o another, in accordance with changes in the circumstances, there is no
tuation in which it does not appear, and urging the subjects to answer
ruthfully only introduces another confusing factor, sometimes causing
ubjects to damage their showing unduly, sometimes apparently arous-
ng defensiveness and causing them to deny characteristics they have
hich might not be damaging in the circumstances. However, response-
ias can be measured and compensations for its effect can be devised.
llis and Conrad are correct in calling attention to its importance, but
re wrong in supposing that it can be disposed of by urging truthful-
ess.

Their third conclusion, that personality inventories "may possibly
e more effective when used with relatively uneducated and less intelli-
ent groups than with groups that are more sophisticated," does not
gree with our experience. A number of psychologically sophisticated
dividuals have taken the Humm-Wadsworth Temperament Scale in
rcumstances where they had considerable stake in the results (as ap-
icants for positions) and some of these have shown unfavorable Pro-
es. Enough of them did poorly to demonstrate that a favorable pic-
re cannot be produced at will. Again, enough of the Profiles of psy-
ologically sophisticated subjects have been verified by case studies
cause us to place more reliance on the Scales furnished by such sub-
cts than we did originally.

The fourth conclusion, that "only limited and specialized demands may be made on the inventory technique," also shows a lack of acquaintance with personality testing. When the data afforded by such measures are examined in a multidimensional fashion from various points of reference, the limitations the reviewers speak of will be largely eliminated. This is not to suggest that a case study should be limited to the use of a personality inventory—in fact, we recommend the opposite; but it is to suggest that the data obtained from such inventories be adequately utilized. Properly interpreted, they can become a valuable aid to the psychological diagnostician and to the psychologically oriented personnel administrator.

Our disappointment with the attempts of the military psychologists to use personality questionnaires for predictions of success in training and in service is not caused by the low correlations yielded by the procedures used. Such procedures must, in the nature of things, have yielded such correlations. Our disappointment arises out of the fact that those procedures and no others were used and that the results they yielded are being taken seriously.

The only one of the so-called experiments reviewed by Ellis and Conrad of which we have personal knowledge is the one concerning the Humm-Wadsworth Temperament Scale as applied to pilots in training. When this project was begun, it was our understanding that it would be carried on by individuals who were familiar with the Scale and that our collaboration would be accepted. When we learned that the work was proceeding along lines that our own experience had already demonstrated to be futile, we protested strongly and finally our participation was again accepted—but, as it turned out, only on the surface. We were furnished the data of age, raw scores, and no-count of 385 subjects, and were requested to categorize them according to suitability or unsuitability of each individual's temperament in such a way that a simple statistical validation against success in flying training could be secured. This we could not do, since the factors entering into such success are complex, and temperament—complex as it is itself—is only one group of these factors. However, we did submit an interpretation of each Scale from the scores sent us and gave an estimate of integration, that is, an estimate of the extent to which the temperamental characteristics found were likely to be under control or out of control. Since some of the temperamental characteristics would be undesirable if out of control, and since some of them would be stimulated to manifest themselves in the training-course situation, but would not necessarily interfere with learning to fly an airplane, a simple correlation of our estimates against pass-fail in the course would be inevitably low.

Again, it was our understanding that we would be permitted to see the data gathered in the experiment and to make our own analyses of them. Whether the conditions under which these data were gathered could have been adequately controlled to provide the conditions called for in an experiment, or whether the cases were or could be followed up long enough or studied in sufficient detail to provide meaningful findings, we do not know. We do know, from our own experiences, that the attempts to oversimplify the problem of ascertaining the contribution of measures of temperament to predictions of success in learning to fly or in adapting to service in the Air Force were doomed to failure from the start. The unsatisfactory results obtained are a reflection, not upon the measures used, but upon the awkwardness of the attempts to use them.

The senior author of this note has long wished an opportunity to state his position with regard to research where one student undertakes a critique of another student's work. His protests against oversimplified and ill-considered studies of the Humm-Wadsworth Temperament Scale have frequently been interpreted as attempts to inhibit such research, and his requests to be informed of studies contemplated and of their progress have been interpreted as attempts to bias the results. This is not, we believe, the case. We believe that scientific research should have a free hand. We ask to be informed of studies contemplated and to be kept informed of their progress only so that such suggestions as we have to make may be given in time to be useful. We do not ask that any findings be suppressed, but only that such comments as we have to make be considered before, rather than after, publication, and that our comments be published with the report of the study.

We realize that this represents a departure from accepted practice, but we believe such a departure would improve the situation with regard to the publication of research. It would remove from an overcrowded literature some futile controversy, leaving room for discussion of fundamental differences; and it would enable the busy reader to consider at one time both sides of such discussions.

Received February 11, 1949.

REPLY TO THE HUMMS' "NOTES ON 'THE VALIDITY OF PERSONALITY INVENTORIES IN MILITARY PRACTICE'"

HERBERT S. CONRAD
Federal Security Agency, Office of Education
AND
ALBERT ELLIS
The New Jersey State Hospital at Greystone Park

The Humms[1] have given an interesting account of their philosophy of personality-test validation. Some elements of this philosophy would, unfortunately, be very difficult to fulfill in a wartime military situation. If, for example, the data from personality questionnaires must always be "examined in a multidimensional fashion from various points of reference," then the usefulness of such questionnaires for wartime military screening—when speed is generally at an extreme premium—must indeed be problematical.

We agree with the Humms regarding the advantages of questionnaires which are broadly applicable to all sorts of subjects (although the question of comparable intelligence, reading ability, motivation, etc. of these subjects would seem to be of some relevance here). It was our conclusion, however, from a review of the literature, that personality questionnaires seem more effective if "specially designed for the group to whom they are applied" (p. 421). The gain in effectiveness achieved by special design seemed sufficiently large to warrant attention. Apparently the Humms' experience fails to confirm this conclusion. Additional empirical evidence is needed on this question.

With regard to the Humms' second main point, concerning honesty of response to personality questionnaires, again we would welcome additional empirical evidence. The Humms are correct in suggesting that merely "urging the subjects to answer truthfully" is not the solution to the problem—nor did we suggest this as the solution. Rather, we pointed out that the *military situation* (not merely the directions accompanying the questionnaires) frequently appeared favorable to reducing dishonesty, and that this fact might account, in part, for the higher validity of personality questionnaires in military practice. The Humms' statement that "response-bias can be measured and compensations for its effect can be devised" strikes a hopeful note—but is only meagerly substantiated as yet.

---

[1] D. G. HUMM and KATHRYN A. HUMM. Notes on "The validity of personality inventories in military practice," by Ellis and Conrad. *Psychol. Bull.*, 1949, **46**, 303–306.

In their third point, the Humms mention that their own experience contradicts the suggestion by Altus and Bell that "more subtle questions [than are contained in most personality inventories] would doubtless be required" for personality-measurement of the more intelligent and sophisticated adult (p. 404). This contradiction in the conclusions of experienced investigators again indicates the need for additional evidence.

As to the fourth point, it seems to us one of the most assured and impregnable conclusions from the literature that "only limited and specialized demands may be made on the inventory technique" for the diagnosis of personality (p. 421). The Humms do not accept this conclusion. Very likely some practitioners succeed better than others in interpreting the information from personality inventories, and perhaps some inventories prove more useful for this purpose than others. Our statement, however, was a general conclusion, based on the experience of various experimenters with many of the most widely used inventories, and not aimed at any particular inventory or individual.

We see no general objection to the Humms' recommendation that the author of a personality inventory be consulted at all stages in an investigation involving the author's inventory. It seems clear, however, that if an inventory were in wide use, the author could not possibly find enough time to answer all inquiries. In the physical sciences and medicine, it is not generally necessary to call upon the author of a new procedure or technique, before the procedure or technique may be properly applied. Some psychologists might argue that such a situation has already been reached in the field of personality measurement by the inventory technique.

Finally, let us remind the reader that the Ellis-Conrad review[2] was not aimed against the Humm-Wadsworth scale nor indeed against any personality inventory. The review attempted, first, to determine "the reasons for the superior showing of the personality inventories in military practice" (p. 385); and second, to draw some general conclusions applicable to the use and improvement of personality inventories in the civilian situation. It is these general conclusions which the Humms, principally on the basis of their experience with the Humm-Wadsworth Temperament Scale, have questioned. They imply that if their specialized techniques of standardization, case history comparison, and interpretation of personality inventories were universally followed, our general conclusions would not hold. The situation calls for additional evidence.

---

[2] A. ELLIS and H. S. CONRAD. The validity of personality inventories in military practice. *Psychol. Bull.*, 1948, **45**, 385–426.

Received March 25, 1949.

BOOK REVIEWS

SECHENOV, I[VAN MICHAILOVICH]. *Selected works.* Moscow: State Publishing House for Biological and Medical Literature, 1935. Pp. xxxvi+291.

This volume was prepared for foreign delegates to the XVth International Physiological Congress in Leningrad and Moscow in 1935. Its purpose is to make known to scientists outside of Russia the nature of Sechenov's contribution to science. The book includes a twenty-eight-page biography of Sechenov, written in English by Sechenov's pupil, M. N. Shaternikov, nine physiological papers on chemistry of the blood and on neural action, printed in the original German and published between 1862 and 1901, and four papers, three of them quite long and all of them of special interest to psychologists, which were originally published in Russian from 1863 on and which were translated into English for this volume.

Sechenov was born in 1829. Moscow was his alma mater. He worked under Glebov at Petersburg (1851). He went to Berlin (1856) to take lectures from Johannes Müller and Magnus, and to do research with du Bois Reymond. At Vienna (1858) he came under the influence of the great physiologist, Carl Ludwig, who remained his lifelong friend and mentor. Eleven of Ludwig's letters are reprinted in this *Life*. Next Sechenov got to Heidelberg (1859) to work under Helmholtz and Bunsen and to form a close friendship with Mendeleyev. After teaching in Petersburg, he worked with Claude Bernard in Paris (1862). He made visits to Munich and Graz (1863). In between times (1860–1870) he taught physiology at the Medico-Surgical Academy in Petersburg. In 1870 he resigned over a controversy about the appointment of Mechnikov and went to Odessa for six years. Then he was elected to the chair of physiology at Petersburg (1876), resigned to go to Moscow as a mere lecturer (1888), was elected to the chair of physiology at Moscow (1891), resigned on account of age (1901) and died in 1905.

The reprinted papers in this volume represent Sechenov's two principal scientific interests. One of his research fields was the absorption of gases by liquids, especially of CO^2 by the blood. Sechenov even invented an improved gas pump which Ludwig immediately adopted in his laboratory. Sechenov's other chief interest was neural action, especially the inhibition of spinal reflexes by the appropriate center in the cortex—the "Sechenov center" it came to be called. American students of physiology used also to hear about the "Sechenov experiment"; you put salt on the upper cut end of the spinal cord and find that you can no longer elicit some of the reflexes that are localized lower down. Sechenov began his publications along both of these lines about 1860, continued the gas problems up to 1875 and the problems of neural inhibition up to 1901.

The first of the Russian papers translated into English is *The re-*

flexes of the brain, published in 1863. It argues we should recognize the dependence of thinking and the other intellectual functions on stimulation, and thus, supporting a mechanistic view of mind, it justifies our naming Sechenov a pioneer in Russian objective psychology. Pavlov himself names this monograph as the inspiration for his work on the conditioned reflexes. Said Sechenov, "All acts of the conscious and unconscious life are reflexes." Movement in a brain reflex may be inhibited as is reflex movement in the cord, but psychical acts, he argued, are clearly initiated by stimulation. Nevertheless Sechenov did not carry his materialism to the bitter end, for he was a valiant defender of freedom of action and the spiritual values. He was a man of deep moral sense, human sympathy and constant concern with causes (like education for women). The Russian intellectuals liked his book, but the Petersburg Censorial Committee in 1866 condemned it as materialistic, forbade its sale and initiated court action against its author for the corruption of morals. The Minister of Justice, however, in 1867 quashed the action and exonerated this enthusiastic, virtuous, beneficent mechanist.

The paper called *Who must investigate the problems of psychology, and how?* was published about 1870 and provides a constitution for the reflexology of Pavlov and Bekhterev. The anonymous editors of this volume say that the schools of Pavlov and Wedensky ("Wedensky inhibition") are based upon Sechenov's views. Bekhterev they do not mention. Both Pavlov and Bekhterev were educated in the Petersburg atmosphere, which Sechenov had created before he went to Moscow in 1888. Certainly Pavlov's views about reflexes, the "psychic juice" (gastric) and psychology within physiology make more sense when viewed against the background of Sechenov.

The last item in this volume is the English translation of a small book which had already had a second edition in Russian. It is called *The elements of thought*, and its eight chapters formulate Sechenov's mature mechanistic views about the nature of intellectual mental processes.

None of these writings of Sechenov's refers by more than casual reference to the work of other authors. Sechenov lectured in the grand tradition of the nineteenth century savant, students swarming to hear and the public eager to read the pronouncements of the great man. He was ready with praise for Helmholtz and other distinguished men of his time, but he did not provide foot-notes referring to their published writings. In spite of his close contact with German and French physiologists and the frequent personal letters from Ludwig (who was first at Vienna and then at Leipzig), the modern reader feels that Petersburg and Moscow were pretty far off from Berlin and Paris. (That was not, of course, true of the Petersburg Academy which entertained so many western savants.) Nor did the editors of this volume provide the means for integrating its content with other scientific work. They do not give the original places and dates of publication and my approximate fixing of dates comes only from the somewhat unsystematic biography that starts the volume off.

We could do well with more knowledge of the history of Russian science. This book, casual as it is in date and detail, helps to give perspective to the later scene in Russian psychology.

Edwin G. Boring.

Harvard University.

Blum, Milton L. *Industrial psychology and its social foundations.* New York: Harper, 1949. Pp. xi+518.

Conservative readers of Blum's *Industrial Psychology* are likely to consider it audacious if not downright revolutionary in its organization and content. Unlike all of his predecessors within the last decade, Blum has emphasized the achievement of human work efficiency more via the vehicle of average employee goals-motivation-and-realization than through the more orthodox vehicles of selection and placement testing and merit rating. Psychological testing is relegated to two chapters in the latter half of the book, and merit rating is completely ignored except for casual mention in a few scattered paragraphs.

This reversal of emphasis—or, perhaps, more accurately, broadening of emphasis—from the traditional approach of Viteles, Moore, and Tiffin immediately marks this volume as either a novelty in its field or a first milestone of a new trend in industrial psychology.

In his preface the author states that much of his material is controversial but he insists that "the social bases of industrial psychology must be understood if the promotion of more harmonious employer-employee relations in a democratic society is to be given realism and meaning." Perhaps his viewpoint is best stated in his introductory chapter.

> If industrial psychology is to be distinguishable from "scientific" management and other systems which may border on its field, it must boldly expound the view that democratic principles must be applied to industry. Its goal should be the satisfaction of man, not of any one man to the disadvantage of others, or of one group over any other. Men must be free to express their feelings, to reach goals, produce, and develop as secure individuals. These and other freedoms are possible in a democracy and are also possible in an industrial system operating in a democratic fashion. Efficiency then follows as a necessary accompaniment (p. 1).

While the reader may wonder how the author proposes to implement such a statement, he is likely to conclude from the content of the book itself that nothing radical is intended. In fact, in the same introductory chapter, he clarifies his conception of the role of the industrial psychologist as being a thoroughly conventional and practical one:

> The psychologist employed in industry has a varied job and varied titles.

> His main function is to render service to the employer, to the employee, and to the various technical specialists in the organization (p. 3).

After a brief review of the work of psychologists in industry, Blum devotes the 36-page second chapter to the "Hawthorne Studies" of the Western Electric Company; these studies he considers as marking the "coming of age" of industrial psychology as an applied science. His next three chapters (84 pages) deal with the measurement, study, and use of the attitudes of employees ("Measuring attitudes in business and industry," "Job satisfaction," and "Industrial morale.") The treatment of attitudes is interesting, provocative, and well documented.

A chapter on incentives is relatively conventional except to emphasize that financial incentives have been much overemphasized to the inadequate utilization of other psychological incentives. Job training is handled well but briefly and a chapter entitled "Time and motion studies" is treated competently in the conventional manner except that this chapter becomes entertaining reading at times when the author pauses to poke fun at some of the short-comings of efficiency studies. Most valuable is his treatment of "breaking resistance" to change not only in introducing more efficient methods but in forwarding other worthy changes.

Blum's treatment of fatigue is not comparable to that of other available psychological treatises, but his chapter on "Work environment" is a valuable contribution. An excellent chapter on accidents is followed by a highly useful chapter on leadership which recommends use of Jenkins' nominating technique in industry. The next two chapters, "Industrial warfare" and "Unemployment," are frank recognitions of stark psychological realities which academic psychologists previously have shown a disposition to dodge or ignore. Blum's viewpoint on industrial warfare is the modern legal one:

> In the present stage of labor-management relations, collective bargaining seems to offer the best solution. If collective bargaining is to be sincere and mature, both parties must have relatively equal power; neither one must have so much as to force the other into submission (p. 341).

Blum is likely to be criticized for not mentioning the problem of responsibility when he discusses power; it seems reasonable that he deserves such criticism. However his resume of the conclusions of Golden and Ruttenberg should serve as a helpful stimulant to inventive thinking in this field, and certainly all industrial psychologists should read this chapter.

The two chapters on testing provide a skimpy treatment by traditional standards, but will be consistent with the current trend to introduce a separate course on personnel testing in the typical psychology curriculum. Highly useful introductions (not present in publications since the 1942 text of Moore) are the last three chapters on consumer research, advertising, and selling. A last chapter on summary and integration tie together some loose topics and provide the student and practitioner alike with a somewhat broader perspective of the problems of industrial psychology than can be obtained from most competing books.

WILLARD A. KERR.

Illinois Institute of Technology.

HOLLINGWORTH, HARRY L. *Psychology and ethics: A study of the sense of obligation.* New York: Ronald Press, 1949. Pp. ix+239. $3.50.

The theme of this provocative and highly readable work appears in its two opening sentences: "The key to morals or ethics is the feeling of obligation, the recognition of imperatives in thought and action. It is the sense of obligation that transforms mere behavior into conduct."

Evidently stimulated by Köhler's treatment of *requiredness*, Hollingworth distinguishes ten primary categories of "oughtness" to which people respond, viz., inference, completion, beauty, welfare, utility, duty, safety, custom, justice, and (somewhat secondarily) legislation. All of these varieties of *should* designated procedures for reducing distress—the notion of the moral, good, or right properly embraces the entire ten meanings.

The author then presents two forms of a 50-item Ethical Insight test scored in accordance with the capacity of the subject to *classify* the propositions in agreement with a key derived from college students and experts; e.g., "This axe ought to have a longer handle" normally reflects the Utility judgment, while "A man ought to tell the truth regardless of consequences" typically illustrates a Duty reaction. The medians increase steadily until early college years; post-Ph. D. specialists barely differ from freshmen in this measure of discrimination.

A novel section rich in clever writing considers the various schools of psychology and their power to explain oughtness satisfactorily. Structuralism and behaviorism are quickly laughed away, but Gestalt theory and the hormic approach are criticized with unusual fullness and rejected as limited principles unable to account fully or adequately for the ten basic categories of "facts." As one would expect from a past president of the American Psychological Association, he offers instead his own familiar triadic system, (1) all motives are distresses, (2) all true learning is "cue reduction," and (3) control is a function of *scope*. This simple pattern is then used to interpret extensively the distinct varieties of obligation sampled by his Ethical Insight Test. Hollingworth does this brilliantly and impressively, notably in his "motor experience" version of the sense of injustice as centered about a demand for a minimum of kinesthetic asymmetry by bilateral organisms like ourselves. The principal emergent suggestion is that optimal ethical acts are those which converge in some "final common path" after weighting all the ten types of ought as these appear in any conflict situation.

No work resting so heavily on the Procrustean technique can fail to raise a host of debatable issues, and this one is peculiarly vulnerable despite the permanent importance of the topics and the admirable ingenuity manifest in their exposition. Either there are *less* than ten basic imperatives in life's affairs or there are *more*; major values in a pluralistic world are probably not confined to this round-numbered schema. Justice, duty, and welfare as "drives" seem phenomenally more alike than different so far as distinguishable pressures behind actual choice decisions and subsequent performance are concerned. One need not be a doctrinaire Gestaltist to hold that coercive or inevitable inference is one manifestation of the completion-tendency in *all* behavior (or vice versa) and that homeostatic tensions to "finish" an indicated total pervade most of the special classificatory principles Hollingworth apparently accepts as irreducibles. His apparent lack of concern with unraveling the complex motivation of any particular imperative proposition opens him to the charge of abstract formalism; his preference for "education rather than social reform" seems like an inconsistent denial of his own oft-affirmed faith in "the natural continuum"; and his claim that socialistic philosophies are psychologically unsound (p. 163) is paradoxical in view of the obviously powerful mass oughts supporting the trend toward democratic socialism in Britain and elsewhere in the contemporary world. More seriously, one wonders about the anti-hierarchical step of making "conscience" merely one ethical category among many when it enters methodologically unrecognized as *the* over-all response "representing action based on full insight and exhibiting full scope" (p. 169).

But these are minor limitations in an exceptionally original, stimulating, and rewarding treatise. Its down-to-earth practical quality should appeal to teachers and counselors generally, while philosophers, social scientists, and psychologists could profitably build many extensive monographs on the numerous research hints hidden in the rich content of this pioneering volume.

GEORGE W. HARTMANN.

Teachers College, Columbia University.

VAUGHAN, WAYLAND F. *Sociol psychology: The science and the art of living together*. New York: Odyssey Press, 1948. Pp. 956.

". . . . I have selected topics which seem to be of primary concern to the student who is seeking a satisfactory personal orientation amidst the social chaos of our time. The central theme is based on the conviction that human beings can achieve happiness in their social relations by discovering just how we've gone about it to make one another so miserable. Surely we can profit by our mistakes (p. vii)." This quotation indicates the tone of the entire volume. It is projected at a homely level and is amply stocked with interesting quotations (between 25 and 30 percent of the volume), a large proportion of them coming from newspapers and the "better" magazines. Its twenty-one chapters deal with such standard topics as "The Individual in Social Situations," "Convention," "Propaganda," "Prejudice," "The Family," and "The Church," and, on the other hand, with such special topics as "Values and the Art of Living," "The Unemployment of Intelligence," and "The Socialization of Intelligence." It is characterized by a willingness to make judgments concerning matters of state, religion and person, most of which are liberal judgments. The book is definitely not academic and is really of a type which the nineteenth century called "moral philosophy," a free-ranging, evaluative (though not necessarily opinionated) presentation of many aspects of inter-personal relationships. Written in a rather personalized style and with a desirable tinge of muck-raking (or is it iconoclasm?), it will make interesting reading for beginning students.

However, the question arises as to whether it is written for beginning students. In most universities social psychology follows at least one introductory course in general psychology, so that some continuity can be expected. While this continuity could be methodological or material in nature, Vaughan avails himself of neither of these possibilities. Instead the text contents itself with popularized explanations not much above the level of those it criticizes. As an illustration, in disposing of Le Bon's theory of the collective mind, the following explanation is offered of crowd behavior "The crowd situation brings out 'human nature in the raw' by releasing our primitive passions. There is no need to resort to the concept of a collective mind to account for our barbarous conduct under such circumstances (p. 620)." This is hardly an acceptable scientific account of crowd behavior though admittedly it is one which is attractive to the layman. Aside from the question of whether Le Bon's theory is not more sophisticated as a conceptual tool, it is disturbing to think that students might attempt to use such an explanation in meeting the complex problems of social life. On the other hand, the volume does contain many suggestions for handling everyday, interpersonal relationships which can only be described as wise and undoubtedly effective.

Assuming that the book was not written for psychologically naïve students, something has obviously gone wrong. There are several possibilities which come to mind. One is that the author believes that there is, in fact, no continuity between general or systematic psychology and social psychology. This is unlikely for he has already written a text in general psychology and throughout the present volume avails himself of the concepts and results of investigators having a distinctly systematic bias, e.g., the Yale and Harvard groups. It is really a question of how the material is used. The difficulty results, in part, from the belief that in order to get across the viewpoint of modern social science it is necessary to use the language and *notions* of the layman, that somehow or other any attempt at systematization will defeat itself. There seems little warrant for such an attitude. To accept the aim of educating the nonprofessional student does not mean that it must be done on his terms, for such an approach can only serve to maintain the very barrier one is attempting to bridge. It is probably the non-systematic orientation of the unsophisticated person which militates against a rational approach to social issues, so that it would seem wiser to insist that the student at least *try* to acquire the conceptual tools of social psychology. Certainly one of the major outcomes of recent research has been to demonstrate how inadequate every day "thought-ways" are for the analysis and remedy of social problems.

But these are, in a sense, peripheral issues, for what we have been discussing are really stylistic decisions; there are more basic considerations which account for the form of the volume. We cannot escape the feeling that this decision to ignore the formal, conceptual processes of science for an emphasis on the "here and now" springs from a strong sense of urgency. Social scientists today are very sensitive to the role which they may and can play as citizens, and there are strong pressures to jump in and do something. This responsibility seems to be the heart of Vaughan's difficulty. He has acutely perceived the problems which face us as members of society, heard the request for answers and has attempted to provide them. Unfortunately, in many instances, different answers would be just as defensible on the basis of present data. What happens then is that personal hunches or guesses are often advanced as if they were *the* answers to the problem which has arisen. The reader has not been adequately prepared to recognize these as only one set of many possible proposals and correctness by authority is thus substituted for a systematic account. Many of the conclusions (we may say) seem correct to us but this simply means that we are both on the side of the angels. Moreover, a scientific account requires more than such mutual agreement, even on conclusions. Thus, while the volume's stress on the need to recognize values and objectives in scientific research is salutary, it commits the error of advancing a specific set of values rather than providing some means of analyzing various normative proposals.

These remarks concerning the need to train analytic rather than convinced readers seem to justify the following concluding statements. It is probably more efficient in the long run to stress the logic of research, the need for more information, the nature of theory and the desirability of caution. It is perhaps more damaging to supply inadequate tools or a false sense of knowledge to students than it is to hold in check a desire to meet one's social responsibilities.

RICHARD A. LITTMAN.

University of Oregon.

PETERSON, H. A., MARZOLF, S. S., AND BAYLEY, NANCY. *Educational psychology.* New York: Macmillan, 1948. Pp. xiii+550.

This new textbook in educational psychology is clearly a superior production, and makes a very useful addition to the literature of the field. It is well written, thoroughly documented, and up-to-date in its orientation and in its use of research material. A number of features of the book are particularly meritorious. These include (1) lists of films at the ends of chapters which should be valuable for instructional purposes; (2) more emphasis upon social learning and group participation than is usually found in such texts; (3) stenographic reports of actual classroom teaching situations which can be studied by the student and evaluated in terms of sound psychological knowledge; and (4) excellent study material and lists of readings in connection with the individual chapters.

The book devotes chapters to all the usual areas of educational psychology, namely, growth and development, learning, personality adjustment, and evaluation. However, the general field of learning receives the major share of attention, with measurement and evaluation probably ranking second in terms of allotted space. In the opinion of the reviewer, hardly enough attention is accorded the general areas of mental hygiene and personality adjustment. The one short chapter dealing with this subject is excellent, but a more extended treatment of classroom problems in this field would probably have added materially to the value of the book. The added material on mental hygiene could perhaps have replaced some of the less valuable discussions of retention in learning, fatigue, diurnal variations in efficiency, etc.

This textbook also seems to suffer somewhat from a weakness which is virtually universal among books in educational psychology. This shortcoming has to do with its partial failure to provide the student with an integrated set of principles or theory which can be used to explain behavior in its many manifestations. A student who masters a book of this type will tend to possess many individual facts, and know the results of numerous experiments, but will scarcely have a unified psychological "frame of reference" which is clearly understood, and which can be applied in the many diverse situations faced by teachers. This criticism, however, is one which can be directed toward most treatises in the general field of psychology. Perhaps the science of psychology has not yet reached the stage of maturity which would make it feasible for it to go very far in setting forth theories or generalizations which would be useful to those engaged in engineering the learning and adjustment process. It would seem, however, that textbook writers of educational psychologies should make every effort in the direction of providing basic theory in those areas of behavior where this is possible.

Peterson's *Educational Psychology* covers in a careful manner much material often treated in general psychology textbooks. It would seem that the general treatment of the subject is such that the book might be particularly useful in those courses in educational psychology which do not require general psychology as a prerequisite. It is, however, clearly a text of such thoroughness and breadth that it undoubtedly will be widely used in more advanced educational psychology classes.

On the whole, Peterson and his collaborators are to be commended for having written a most readable, practical, and authoritative text which will exert a very wholesome influence upon teacher training programs.

GLENN M. BLAIR.

University of Illinois.

BRAY, CHARLES W. *Psychology and military proficiency.* Princeton: Princeton Univ. Press, 1948. Pp. xviii+242.

This book should be required reading for Army, Navy, and Air Force officers. In summarizing the history of the Applied Psychology Panel of the National Defense Research Committee, Professor Bray has done a service for psychology and for the military. He shows how it came about that civilian psychologists were enabled to render notable aid in the solving of many of the human problems of the military during World War II. Many of these problems were completely new to psychology and it is a remarkable tribute to the ingenuity of the Panel's investigators thay they succeeded in adapting the techniques of experimental psychology to this use.

Professor Bray brings order to what has been a confusing picture. He delineates clearly the relationship of the Applied Psychology Panel to other wartime research agencies, to the Panel's 23 contractors, and to the various military agencies.

The results of the Applied Psychology Panel's work have been reported in 513 separate documents. Too frequently the author is forced to note that a particular report is "not available to the public." While it is understandable that some of the investigations may possibly continue to involve national security, it is to be hoped that declassification is not being delayed only because of lethargy on the part of the responsible military agents. While Bray's descriptions of the various research projects are necessarily brief, he does give the reader enough of a taste to whet the appetite. To find that the original research report is not available is most frustrating.

The program of the Applied Psychology Panel was carefully planned, planned in such a way that optimal use would be made of the limited personnel resources available to the Panel. This personnel limitation was more serious than is generally known. A count of the people mentioned by Bray as having been engaged in Panel work discloses that fewer than 200 individuals were involved. Because of this limitation, the Panel was forced to reject many service requests for help. Early in the work, the major emphasis was on classification and training. This emphasis shifted gradually until, at war's end, a large portion of the Panel's resources was concentrated on research on design and operation of equipment and on achievement and proficiency tests.

Professor Bray, in the final chapter, "The Future of Military Psychology," makes perfectly clear what the role is that psychology should continue to play in the military services. He summarizes most effectively the conditions under which psychology would be able to make its greatest contribution to the military and he challenges the military to provide those conditions.

GLEN FINCH.

Headquarters, U. S. Air Force.

EYSENCK, H. J. *Dimensions of personality*. New York: Macmillan, 1948. Pp. xi+291. $5.00.

In our present state of knowledge, any large-scale investigation of personality is an important event. This book is an account of the work that has been done at Mill Hill Emergency Hospital by a number of psychiatrists and psychologists working together. The institution is one which served chiefly men and women who because of neurotic characteristics were unable to adjust to army life. Since most of the specific studies have already been reported in journals, Eysenck does not attempt to describe them all in detail but tries rather to bring them together into a coherent organization and fit them into the context of past research on personality. He does this well, analyzing probable reasons for discrepancies from study to study. His definitions of terms and concepts, clarifying what he and his co-workers mean by trait, factor, introversion, etc., are very helpful in this connection.

As the title indicates, the first task was to identify the principal dimensions along which personality might be measured. For this purpose, 700 patients were rated by psychiatrists on 39 traits. Factor analysis of correlations between ratings identified a general factor they called P, general personality organization, or general neuroticism. The second factor, bipolar in nature, differentiated anxious, inhibited neurotics from hysterical, asocial ones. They called it dysthymia-hysteria or introversion-extraversion (in the Jungian rather than the Freudian sense). These two factors together accounted for 26 per cent of the variance. Two other bipolar factors, representing eight per cent and six per cent of the variance respectively, were considered to be of little fundamental importance.

From here on the research activity consisted of an exploration of the ways in which the two basic dimensions are related to a wide variety of other factors. A very catholic attitude seems to have characterized their activity. Physical and physiological measurements, ability tests and personality questionnaires, laboratory performances of several kinds, and projective or expressive techniques were all tried.

A few strikingly significant relationships emerged. General neuroticism seems to be related to a significant extent to poor scores on dark vision tests and to body-sway measures of primary suggestibility. Dysthymics are clearly differentiated from hysterics in some aspects of their level of aspiration scores. There are many other *suggestive* findings and *promising* techniques, such as the Ranking Rorschach Test and the analysis of responses to humor of various sorts. In the last chapter, the author brings together the bits of information into a composite picture, but there are really not enough fairly certain facts to make this very convincing.

The weaknesses in the work are those which would be likely so show up in a situation where many individual investigators branch out into separate paths without much central control over formulation of problems and methods. Although it is stated that 10,000 subjects were involved in the whole undertaking, the number of cases in many individual studies is entirely inadequate to settle the issues being investigated. Ways of showing relationship between the basic dimensions and other traits vary from one study to another, and many of them present data in an awkward form which makes interpretation difficult and comparison with other studies impossible. One wonders why the weighted scores for neuroticism and introversion which were used in plotting Figs. 4 and 5 could not have been used in all the separate studies and correlated with measurements of the traits being investigated.

The most serious limitation of the work arises from the non-representativeness of the population, especially that upon which the initial factor analysis is based. Patients at this hopsital were not even representative of neurotics in general, and men and women appear to have been differently selected. The author, of course, recognizes this difficulty, and it is to his credit that he describes the experimental group in some detail and brings in comparison data from normal populations wherever possible. These comparison groups, however, are very diverse and often very small (e.g., twenty normal controls were used in the exercise response study, p. 103). Sometimes no information is given as to how they were chosen (e.g., Ranking Rorschach study, p. 212).

In a largely uncharted area like personality research it would be unfair to judge a preliminary voyage of exploration by ideal standards. Limited data are certainly of a great deal more value than no data at all. The main task accomplished, as the author sums it up, is that of "isolating two personality dimensions and discovering a series of tests which enable us to perform quantitative investigations along those two dimensions." This alone is a considerable accomplishment. It is as a framework for further quantitative personality study and as a rich mine of ideas for laboratory and test methods by which one might proceed that the book will be most useful. It is in no sense the last word on the subjects it discusses.

LEONA E. TYLER.

University of Oregon.

WHITE, ROBERT W. *The abnormal personality, a textbook.* New York: Ronald Press, 1948. Pp. x+613. $5.00.

In *The Abnormal Personality,* Robert W. White contributes enormously to the need in teaching for an integration of personality theory with the traditional precepts of psychopathology. It is apparent that the author is well grounded in sound, experimental psychology, and experienced in actual clinical practice, and that for some time he has been preoccupied with the problem of an amalgamation of these in the form of a theoretical structure which would be useful in teaching. Perhaps more important than these characteristics, however, are two which the reviewer would like especially to emphasize. Manifest throughout the book, these are qualities which, in combination, seem today to be exceedingly rare: scholarship and good writing. This reviewer would like to discuss each in turn.

The reader of *The Abnormal Personality* is aware soon enough that White's position is very clear with reference to much of the theoretical controversy regarding personality, involving as it does psychoanalysis, Kraepelinian psychiatry, present-day emphasis on clinical psychology, "organic" vs. functional etiology, psychosomatic medicine, etc. He seems to have assumed it as his special task to disentangle from this complex the threads of significance, with the possibility that these might form a pattern which the student in the course in "Abnormal" might assimilate. In doing this, White has been almost compulsive in his use of basic literature. It is apparent that the author has read widely. But his writing indicates that he has read not only widely but carefully, and that he has been selective in his references. His study has been intensive as well as extensive. Because of this feature alone, one might attribute scholarship to his book. But a factor which counterbalances and yet supports this is the author's willingness, when facts are scarce, to put his best foot forward, humbly but with common sense, and with the other foot firmly on the ground. It seems to the reviewer that we have here an exemplification of psychological study—examination of research findings, combination with personal experience, incorporation with most reasonable current theory—at its best.

In regard to the matter of good writing, this is a book which goes easily along from its first to its last page. The author seems to have made a real effort to use ordinary language to establish basic concepts which the old-fashioned jargon of traditional psychopathology invariably confuses with its Hellenicized mumbojumbo. It would seem that, to the author, clear speech and, what is most important, clear thinking are of paramount consideration.

In certain areas the reviewer is not in agreement with the author's position. (1) With a healthy orientation in regard to psychodynamic probabilities, White seems to take Sheldon pretty seriously and to keep a seat reserved for the geneticists if they should happen to turn up with anything. (2) Goldstein's position with reference to "organic" possibilities is well presented provisionally, but it seems to have been adopted generally with reserve. (3) White is at his best when discussing neurosis and its dynamics; regarding psychosis he seems to take much more refuge in the authorities.

Essentially, however, this is an approach to the subject-matter of "abnormal psychology" on the one hand, and "personality" on the other—areas traditionally disparate, but actually synonymous—for which the reviewer, in particular, has been looking for a long time.

T. W. RICHARDS.

Northwestern University.

MACKINNON, DONALD W., AND HENLE, MARY. *Experimental studies in psychodynamics: A laboratory manual.* Cambridge: Harvard Univ. Press, 1948. Pp. ix+177.

The disparity between the amount of speculation in the field of human motivation and personality, and the quantity of experimental techniques and evidence in this same area has been disconcerting for too many years. Therefore it is rather pleasent to note the advent of a laboratory manual which has as its modest purpose the presentation of 14 experimental designs to fulfill the laboratory needs of a semester course in the experimental psychology of personality. Such a course, no matter how limited the implications of the experiments may be, will have the highly redeeming feature of focussing the attention of students of personality upon experimental problems and will force them to think in terms of experimental design. In the ability of this manual to contribute to such a course lies its chief merit.

In addition to its mundane pedagogical aspects, this book has certain theoretical implications. If one were to accept uncritically the panegyrical descriptions with which some theorists characterize their theories, one might be seriously misled as to the degree of maturity the science of psychology has achieved. Fortunately, we can examine what psychologists do in addition to what they say they do. Since the authors of the experimental manual are kindly disposed towards Lewin's theory, the experiments, many of which were initially reported by students of Lewin in German periodicals, are considered within the framework of topological and vector psychology and consequently an opportunity to examine Lewin's theory on the experimental level is gained.

A rather interesting and significant methodological point that is repeatedly raised by the authors is that the experimental behavior must "be understood in terms of *the psychological field* in which it occurred" and that "it is insufficient to be able to specify the characteristics of the physical situation in which the subject is placed and to correlate these with the experience and behavior of the subject." They admit that it is a rather difficult task to determine the subject's psychological field but nevertheless assert that it is possible. Clues as to the nature of the subject's psychological field can be gained from the observation of the subject's behavior and his verbal reports. But "these clues must be gained subtly and examined critically." One wonders about the fruitfulness of such an evasive construct. There is nothing methodologically incorrect in interpreting theoretical constructs as reflecting subjective, conscious experience, but by so doing one does not escape the requirements of objective, operational usage of concepts. It has appeared to the reviewer that the system of constructs used by Drs. MacKinnon and Henle has been somewhat more phenomenologically descriptive than theoretically and empirically consistent.

The manual was used by the authors with students who had some experience with experimental procedures in psychology. One might assume that training in clinical psychology would be advisable since in one experiment "a subject should be chosen who is lively, active and energetic, one who takes pride in what he does, who likes to do things well" while another experiment requires a subject "who is sensitive and creative, though not at all a prima donna."

Each experiment is designed for a four-hour laboratory period. With the exception of a few simple pieces of apparatus the laboratory material for the entire course is supplied with the manual. A booklet for the instructor has been prepared to provide helpful suggestions. For those who find themselves in agreement with "the field" approach, this manual can serve as the core of an experimental course in "psychodynamics." The manual should be of interest to anyone who teaches a course in general experimental psychology. Several of the experiments can be used to advantage in such a course.

Howard H. Kendler.

New York University.

Ewen, John, H. *Mental health.* Baltimore: Williams & Wilkins, 1947. Pp. 270.

This volume, written by an English psychiatrist, is intended as "a practical guide to the disorders of the mind." Its title *Mental Health* is quite misleading since it deals neither with principles governing the normal personality nor, to any great extent, with problems of mental hygiene. As a matter of fact, the author has given us, in essence, another volume of familiar descriptive psychiatry. In the light of the ever growing array of publications in the fields of mental disease and behavior abnormalities the reader may justifiably ask, when confronted with a new work, what is being said here that has not been said before or what *new* viewpoints and concepts are being introduced? The present volume meets both questions in a very unsatisfactory fashion.

The author has intended this book to serve as a simple compendium and introduction to the field of psychological medicine. In the opinion of this reviewer, however, his only achievement has been in terms of brevity. In 270 pages he has not only dealt with the aetiology, symptomatology, classification and description of the various mental disorders, but also managed to include a rather lengthy chapter dealing with Legal Acts governing mental disease in England as well as various certification procedures applied to medical officers in mental hospitals. Regarding the intended simplicity of conveying the subject matter, the author has failed conspicuously. Particularly in the discussions of general concepts the beginner will be confronted by a mass of highly ambiguous terms and concepts, undefined and stated in an uncritical and regrettably dogmatic form. The more sophisticated reader will be disturbed by unfortunate inaccuracies as well as by the author's rather archaic definition of "mind" and his apparent obliviousness of more modern concepts of personality mechanisms. The short glossary of technical terms is rather inadequate and shows a poor selection.

While certain readers may find some use for the chapters dealing with the legal aspects of mental disease in England, the beginner interested in being introduced to the field of mental illness is likely to become more confused than educated.

LUDWIG IMMERGLUCK.

University of California, Berkeley.

MENDEL, ALFRED, O. *Personality in handwriting.* New York: Stephen Daye Press, 1947. Pp. 375.

Rudolf Arnheim, introducing Mendel's book, presents graphology as a projective technique, while the author himself calls it "the psychology of man's finest expressive gestures." In the opinion of the present reviewer projective and expressive methods should be differentiated from each other. Projective methods (Rorschach, TAT, doll-play etc.) diagnose a subject's projections of ideas upon a given, more or less structured material, while the analysis of gestures or expressive movements deals with a direct motor expression which is neither projected upon structured material nor transformed into any content.

While many non-graphologists explain graphic gestures by learning and by chance, Mendel starts with the assumption that graphic movements are patterned by personality. "Every gesture may fuse three currents in man: the emotional, the intellectual, and the physical. Likewise every stroke of his pen." However, with the admission, "I have speculated freely, drawing tones and shades from the slightest clues," Mendel plays into the hands of those who condemn graphology and does not do enough justice to himself. He relies in many ways on previously tested relationships between certain graphic movements and certain personality trends and between certain movement configurations and personality patterns. The lack of experimental data, of methodology, and of a bibliography is partly offset by detailed tables of interpretation and by the use of graphology in diagnosing normal and abnormal aspects of personality, supported by empirical observation.

Especially interesting is the chapter done on "Physiology and Pathology in Handwriting" by Alfred Kanfer, who validated his graphological observations of patients by hospital diagnoses.

Mendel follows the statement of Gardner Murphy, which he quotes in the beginning of his book, namely, "not to maintain that the interpretation of character by handwriting is a science; its most devoted adherents would make no such claims." As far as graphological diagnoses can be validated by empirical experience, as far as successful predictions can be made far above chance expectancy, as far as graphic movements can be measured in their pattern and evaluated in their expressiveness, as far as the methodology is consistent, graphology has to demand a scientific recognition. In the opinion of the present reviewer, the direct analysis of expressive movement without intermediary material for projection potentially allows a more precise and more defined application of scientific criteria than projective methods. Mendel's illustrations and interpretations enrich our dictionary about the language of personality but they do not reveal the laws of its grammatical structure, those laws which govern graphic movements as one of the many expressions of personality. When Mendel states, "I hope the experimental psychologist will take my hunches, examine them, then reject or verify and refine them," the present reviewer is one who supports this appeal.

WERNER WOLFF.

Bard College.

GESELL, ARNOLD, AND AMATRUDA, CATHERINE S. *Developmental diagnosis.* (2nd Ed.) New York: Hoeber, 1947. Pp. xvi+486. $7.50.

The revisions of, and additions to, the first edition of this book hardly justify publication as a new edition. Eleven of the 19 chapters of the first book remain completely unchanged. One chapter, called "Developmental Pediatrics," has been added to the new book. The first edition chapter, entitled "Special Sensory Handicaps," has been divided into chapters on "Deafness" and "Blindness" for the new book, and illustrative cases have been added. The only other changes in these chapters entail the addition of a section on aphasia, rewriting of one paragraph in one chapter, and the addition of two paragraphs to each new chapter. Other changes in the book of any consequence are the addition of material on electro-encephalography in diagnosis of convulsive disorders, a brief section on visual cerebral injury, and inclusion of a small amount of material from another recent book of the senior author, "*The Embryology of Behavior*." Additions to the appendices include a new list of Dr. Gesell's films, a very brief list of child guidance books (three of which do not include Dr. Gesell's name as an author), and a short appendix on professional training for developmental pediatrics.

As in the first edition, the material is presented in a chatty, informal manner with only rare references to the experimental or clinical literature. Many vague, unscientific concepts expounded in the first edition persist unchanged in the second edition. The book no doubt has some value for the practitioner, but it offers little to the psychologist or child development specialist.

MARGARET R. KUENNE.

University of Wisconsin.

TYLER, LEONA E. *The psychology of human differences.* New York: D. Appleton-Century Co. Inc., 1947. Pp. xiii+420.

The psychologist who would write a text on differential psychology encounters a number of problems. The appraisal of the text can be based in part upon the way in which these problems are faced. The author must decide: (a) whether to strive for a broad sampling of topics or for thorough discussion of particular problems; (b) how to relate the experimental findings to the controversies and the diverse points of view in the field; (c) how to approach the statistical problems which underlie any treatment of differences between individuals.

In her book, Dr. Tyler has resolved the first problem in favor of a broad coverage of a variety of topics. She undertakes to discuss both group differences and individual differences. She presents chapters both on the feeble-minded and on factor analysis. She treats both the heredity-environment issue and the use of test batteries for the prediction of vocational success. Since the total book is only 400 pages, the treatment of many topics is suggestive rather than exhaustive. At the same time, however, the book provides a good introduction to a wide range of problems relating to individual differences between persons.

In her presentation, Dr. Tyler tries to distinguish between the facts which have been gathered for a particular area and the interpretations which are to be made of them. Her point of view, to which the reviewer would subscribe, is that psychologists differ very much more with respect to what the facts of individual differences *mean* than with respect to what the facts *are*. Each chapter undertakes to present a selection of the most significant facts in a field, and then to discuss the problems of interpreting those facts. The evidence selected for presentation seems generally well chosen, and the interpretation of the facts seems to present a mature point of view which avoids doctrinaire positions on either extreme. Maybe this merely means that the point of view is in most cases one with which the reviewer finds himself in agreement.

What to do about statistics is always a difficult problem for this type of a book. One has the choices of (a) writing a book for which no statistical competence is needed, (b) assuming that each reader already has competence in statistics, or (c) trying to include instruction in statistics in the text on individual differences. The first choice necessarily results in a book which is superficial to the point of triviality. The second choice is probably unrealistic for many groups and institutions which would want to use the book. The third choice puts a distressing responsibility for a dual function upon the author. A chapter of statistical instruction often seems not to fit well into the book as a whole, and at the same time provides a very unsatisfying course in statistics. Dr. Tyler has adopted a variant of the third technique, but the statistical material is quite limited in amount and makes no pretension of developing competent computers. It is slipped in rather unobtrusively at various points in the book where it is needed, and would appear to be fairly painless. However, one may question whether it provides sufficient background for the student to read with very great understanding the sections of the book dealing with use of tests for prediction and with the effort to develop pure tests of primary abilities. Perhaps no better solution is possible for this very troublesome problem.

In general, Dr. Tyler has written a clear, readable, succinct book, presenting an overview of many of the problems of individual differences and a sound appraisal of the difficulties both in conducting the research and in interpreting the results. The text should be quite useable both for advanced undergraduate students and for graduate students whose field of specialization lies outside the limits of differential psychology.

ROBERT L. THORNDIKE.

Teachers College, Columbia University.

ALSCHULER, ROSE H., AND HATTWICK, LA BERTA W. *Painting and personality: A study of young children.* Chicago: University of Chicago Press, 1947. 2 vols. Pp. xi+590. $10.00.

Painting and personality is an important landmark in the study of personality and in the study of early childhood. It is equally noteworthy in the annals of publishing and printing techniques: seldom has a publisher contributed so greatly to the presentation of research as has University of Chicago Press in the 120 magnificent full-color plates by a new process.

Beginning in 1937, when such an approach was far from commonplace, Alschuler and Hattwick undertook to study through individual case studies and quantitative group comparisons the dynamic relationships between inner personality tendencies and spontaneous creative behavior in 2–5-year-olds. With regard to the general expressive hypothesis, Alschuler and Hattwick confirmed the expectation of individuality in each child's "characteristic and unique abstract use of color, line, form, and space" (p. 3). But beyond this their data revealed to them "a very general, and perhaps even universal, tendency for all sorts of individuals to express similar feelings, reactions, and problems in like, or at least comparable, fashion" (p. 4).

These findings are based on observations and daily records of 149 children over one or two years from various social settings. The *individual* analytic procedure began with the finished paintings. Each was analyzed in terms of use of mass, line, form, space, color, recognizable content, verbal comments. Then, behavior and personality data for the same day were examined and an attempt made to note relationships or lack of them. Appearance of consistent trends in various children produced hypotheses for later quantitative study. The forms for breaking down the daily qualitative records of behavior in creative and free play situations and the charts and summaries for compiling the behavior data over extended periods may well be useful to other workers in the field (pp. 224–259). One case study is given in extended summary form.

Group analysis was carried out by determining which children preferred each research medium; and within the easel painting medium, which children showed each of some fifty-odd characteristics (e.g.: Emphasis on Red, Emphasis on Cold Colors; Intermingling of Colors; Form vs. Color Emphasis; Diagonal Strokes; Dotting; Painting Off the Page, etc.). Children showing a certain characteristic were compared with the group as a whole or with children showing a contrasting emphasis in their paintings on a series of 113 personality characteristics under 21 general headings (pp. 235–237). Tables and interpretive summaries present those relationships (pp. 381–582). "*Only when the chances were 85 or more in 100 that the observed tendencies would be found in a similar sampling, was a given relationship included in a table*" (p. 381—italics in original).

It seems to this reviewer, however, that still greater caution is necessary in statistical terms. Admitting relationships at a level of "85 chances in 100" means, of course, reliance on critical ratios barely over 1.0. Moreover, some of these relationships apply to a *minority* of the children in a group. For example, among the traits found to be characteristic of children emphasizing red in their palettes is one labeled "dependent on adults for affection." Examination of the appropriate table (p. 387) shows this trait was found in 39% of the 46 children showing emphasis on red. Among the first 300 of several thousand critical ratios presented in the tables the reviewer counted only 4 of 3.0 or more.

The major findings of the study, based apparently on both individual and group quantitative analyses appear in the first 165 pages of the work. The major conclusions (one would surmise: the structural plan of the book is not always clear) appear to be those stated on page 14. They are in fundamental agreement with Rorschach's analysis of perceptual organization and are worth quoting in full:

> a) Color tends to give the clearest clues as to the nature and degree of intensity of the child's emotional life.
>
> b) Line and form tend to give the most intelligible clues to (1) the amount of energy the child is expending, (2) the degree of control the child is exercising, (3) the direction in which that control is operating.

> Color and line and form considered together are likely to indicate the balance which exists between the child's impulsive drives, on the one hand, and his overt, controlled behavior, on the other hand.
>
> c) Space usage and spatial pattern tend to give less a picture of the child's inner life than a picture of the child as he relates, and is reacting, to his environment.

The elaboration of these factors follows in rich detail, supported by references to the quantitative tables already described and—rather more impressively—by descriptions of the behavior of children showing markedly the painting-characteristics in question. These materials will provide the starting point for new research in this field for years to come, and will undoubtedly be of extreme value in the clinical interpretation of children's painting. But most disturbing here and throughout the study is a kind of ambivalence. On the one hand, e.g., the personality characteristics of children preferring red, blue, green, yellow etc., respectively, are outlined. On the other hand, the authors find it necessary to emphasize that these are dangerous generalizations, not intended for sweeping application. Each painting characteristic has a *potential* significance but it is "more often in the interrelationships of several aspects than in any single characteristic that the distinctive and telling qualities of the child's products are likely to be." But, if this is so, one wonders why so relatively enormous a preponderance of the research itself and of its publication are devoted to the thin gruel of the admittedly inappropriate study of the relation of single painting traits to single personality traits, particularly as the admirably complete presentation of the results shows them to have such disappointingly low statistical significance. Nor is there sufficient guidance to those who would follow in how to weigh the "interrelations of trends"—for it is seductively easy to take separate factors and characterize the child who exhibits them, even though one is warned against doing so. This ambivalence is not simply that of the authors but, to this reviewer, reflects rather clearly the state of psychological investigation of personality, in general. We are torn between the obvious demands of the personality structure itself that it be viewed holistically, and at the same time most of the research techniques available force us to an extreme particularism. Alschuler and Hattwick have done yeoman service toward the development of each horn of this dilemma, but they have not helped us to get from one horn to the other. The pervasive flaw of the work, monumental as it is, lies in the assumption that each aspect of painting must be "*of necessity*, studied by itself without regard to the way in which that aspect of painting might be related to other aspects" (p. 5, reviewer's italics). The next major step in personality research will come as methodology removes that necessity.

LAWRENCE JOSEPH STONE.

Vassar College.

PENNINGTON, L. A., & BERG, I. A. (EDS.). *An Introduction to clinical psychology*. New York: Ronald Press, 1948. Pp. xv+595.

This is a multi-authored volume which is unique in that the editors have not written any of the chapters, and indeed, it is difficult to ascertain what role it was they performed in producing this book. For it is this reviewer's impression that we have here a decidely uneven book, with a modicum of editing and chapters which vary widely in quality.

Although the editors state that the book is "to describe eclectically at the introductory level what clinical psychologists do," we need to know what is meant by the introductory level. If it is the first year of graduate study, much of the book is repetitious of materials taught in competent undergraduate courses. Is the introductory level, on the other hand, represented by an undergraduate course, unless it is deliberately organized around a book such as this? Perhaps the chief contribution of this text may be in smaller institutions faced with the problem of teaching an undergraduate course in "clinical" psychology.

The book is divided into five parts—Introduction; Clinical Problems; Clinical Methods; Psychotherapy; and Professional Relationships. While the editors should be given credit for an awareness of the fact that an orientation to psychopathology is necessary for the introductory student, there is a logical, rather than a psychological, breakdown among these various parts. Thus, the *content* of abnormal psychology (Part II) is presented as distinct from the *methods* of clinical psychology (Parts III and IV). The section on methods, for example, becomes largely a matter of straightforward description of techniques without much discussion of the way these techniques operate except in a few case histories. A review of the various chapters may give a clearer concept of the pattern of the book, and point up its unevenness.

Part I, "Introduction," opens with Chapter 1 on the "Meaning of Clinical Psychology" (R. B. Cattell). While the training situation is treated honestly and forcefully, some may take issue with the author's statement that clinical training applicants are those who cannot face the "discipline of the natural sciences." There are others, too, who would doubt that there is more future for clinical psychology through factor analysis than through more dynamic or intuitive approaches.

Chapter 2 is a discussion by O. H. Mowrer on "What Is Normal Behavior?" Here it seems that the editors could have eliminated the discussion of statistical normality, psychographs, invalidity of pencil-and-paper tests, etc., which is probably repetitious for the student. The chapter is cleverly written as a symposium of specialists in various areas concerning the problem of normality, but the attempt to interject the viewpoints of the different specialists gives a kaleidoscopic feeling which perhaps proper editing might have reduced.

Part II on "Clinical Problems" opens with Chapter 3 by D. W. Dysinger on "Signs of Personality Disintegration." His stress on viewing the individual as a dynamic whole is a precaution well stated to the beginner. The illustrative case material, half of the chapter, is well presented. However, there is a very important point of view which the author apparently overlooked in his discussion of the clinician's responsibilities. This is the recognition of the clinician's own problems, and his awareness of his means of handling them—a point to be stressed early in training.

Chapter 4, concerned with "Behavior Problems of Children" is written by G. L. and H. H. Anderson. What is particularly one-sided is the emphasis upon dominative behavior of parents as a prime source of maladjustment, and one wonders why there is utter neglect of the dynamics of rejection, sibling rivalry, over-protection and submission. Domination is treated from the standpoint of the objective effects upon the child and his overt resistances to it, but no mention is made of the psychodynamics involved. (Neither Symonds nor Levy is mentioned in the bibliography.) It would seem that the editors, as clinicians themselves, should have been impressed with the one-sidedness of this chapter.

Mandel Sherman, in Chapter 5 ("The Mental Defective"), confines himself to a discussion of syndromes encountered in mental deficiency—cretinism, mongolism, etc. A paragraph of some twenty lines devoted to environmental deficiencies, in the opinion of this reviewer, does violence to this very important topic by nature of its very brevity. There is also no treatment of the personality structure of the defective, especially as delineated by Kounin.

Chapter 6, "The Physically Handicapped," by S. Marzolf, discusses the sensory, motor, and other disabilities (chorea, encephalitis, etc.), from the standpoint of etiology, description, and diagnosis. While very good from these standpoints, there is very little discussion on the meaning of physical handicap to the patient, although reference is made to the recent work of Barker, Wright, and Gonick in the bibliography. Again, it is felt that stress should be not so much upon static symptoms and syndromes but upon dynamics of the situation.

R. M. Bear's Chapter 7 on "The Educationally Backward" is a generally adequate treatment of disabilities in the school subjects. A minor error occurs on page 143 where the Binet is listed as a test which does not place a premium on verbal capacity, an error which could have hardly escaped the editorial eye.

Chapter 8 on "Speech and Personality," by F. H. Sanford, is somewhat of a condensation of his 1942 article in this journal. There is thus a difference in style and approach between this chapter and the others, concerning which the editors are aware. There is much more of a research orientation (as witnessed by 74 references, more than any other chapter, and by the content). A debatable statement occurs at one point (p. 174), where it is stated that "there have been no major changes in the theory of aphasia and of allied disorders since the days of Head," with reference to the latter's 1926 publication. Thus the work of Kurt Goldstein and his associates in this area is neither discussed nor cited.

A. C. Washburn's treatment of "Tension States" (Chapter 9) has particular value in orienting the student to the stresses which bring about these conditions. The discussion of even the physiological and neurological aspects is geared at a level thoroughly comprehensible to the beginning student. However, one would like to have seen more case material drawn from the rich psychosomatic literature, since the latter is referred to in the bibliography.

"Psychosexual Behavior" is the title of Chapter 10 by J. Ball. Basically, there is a good succinct presentation of the analytical theory of psychosexual development as a background. One point is moot, though, namely the fact that Ball states that the latency period has little significance for later sexual problems. However, it is the repressions of latency which create difficulties in adolescence. Her conclusions on unhappiness as related to extramarital relations may be too heavily weighted with results from studies of clinic cases.

Chapter 11 by R. G. Kuhlen covers "Psychological Trends and Problems in Later Maturity." It covers the field well, emphasizing the increase in numbers of the aged, change in abilities and adjustments, and practical aspects of the problem. Unlike other chapters, however, there is no illustrative material whatsoever on the senile personality, and thus there is editorial inconsistency.

Chapter 12 by H. Cleckley on "Antisocial Personalities" is perhaps the weakest chapter of the book. There are three principal sections on the delinquent, the criminal, and the psychopath. The section on the delinquent is merely a series of brief descriptions of delinquent personalities, with very little on the causal factors in delinquency. Although the work of Healy and Bronner and of the Gluecks, e.g., is noted in the bibliography, the author gives no account at all of the findings of these writers in this important field. The bibliography does not cite any sociological work by people like Shaw and McKay.

R. M. Dorcus's treatment of "The Psychoses and Psychoneuroses" constitutes Chapter 13, and is essentially an overview of any abnormal psychology text. Case histories, especially for the beginner, are particularly in need of editing. One wonders what the following medical terminology, e.g., means to the student without medical orientation (p. 280): "epithelioma of the bridge of the nose and senile keratosis below the outer canthus of the left eye . . . abdomen distended and tender but not fluid percussed." Emphasis of discussion is uneven—27 pages devoted to organic psychoses; 13 to functional psychoses; and only 8 to psychoneuroses. Although Dorcus is aware of the emphasis upon the organic, it is certainly at variance with his own statement, on p. 313, that "psychoneurotic disorders represent the major area in which the clinical psychologist is most likely to function."

Part III on "Clinical Methods" consists of four chapters. Chapters 14 (J. E. Horrocks, "The Measurement of Achievement and Aptitude") and Chapter 16 (W. T. Donahue, "Interest and Personality Tests") are descriptions of the more commonly used tests of achievement, aptitude, interest, and personality, largely of the paper-and-pencil type with discussion of concomitant problems related to testing and some illustrative cases in which tests are used. A. Magaret on "Intelligence Testing, Clinical Practice" (Chapter 15) has an excellent discussion of the uses and limitations of the individual tests of intelligence in the clinical situation. For the beginning student the presentations of the difficulties inherent in current "scatter analysis," especially with the Wechsler-Bellevue, and of observation of client behavior in the test situation are particularly commendable. The discussion of the Shipley-Hartford CQ Scale, however, fails to indicate its demonstrated invalidity. H. Sargent's Chapter 17 on "Projective Methods" is effective in introducing the student in lucid style to the concept of projection and to the tests themselves. The discussion on the theoretical background of projection and the sections on cautions in use of the projective methods, especially her emphasis on the variable of the clinician himself, are particularly effective.

Part IV, "Psychotherapy," opens with Chapter 18 by F. McKinney on "Directive Techniques." It is essentially a review of the directive processes, especially as used in the college counseling situation. There may be some misleading connotations, though, in that techniques common to practically all therapy appear to be subsumed under the directive heading. For example, on p. 458, under "Forms of Directive Therapy" appears "Expressive Therapy" with brief discussion of catharsis; on p. 450, under "Directive Therapy—The Counseling Atmosphere" is a short description of Rogers' permissivesness! Group psychotherapy is also mentioned as a directive method, which it may or may not be.

"Client-Centered Therapy" is Chapter 19 by W. U. Snyder. There are statements here which the critical beginning student may find bewildering. The directive approach is termed "negativistic." Again, on p. 470, he states that "a unique theoretical contribution to non-directive therapy which definitely affects the technique is Rogers' belief that the use of psychometric tests can interfere with the process of therapy" and only six pages later (p. 476), in presenting a case, Snyder states, "at this interview the results of the recent vocational and personality tests were discussed with Mr. D. . . . " This is followed on p. 477 with a whole page of test results! Of particular value, however, is Snyder's frank discussion of the limited value of the non-directive method with certain types of clients.

Chapter 20, "The Semantics of Maladjustment" by W. Johnson, while presenting a good theoretical treatment of this problem, suffers by contrast with the two preceding chapters in not illustrating just how the semantic therapist operates. The specific techniques of the semanticist are couched in general terms, apart from mention of procedures common to all therapies. There is complete absence of illustrative case material, a point upon which perhaps the editors should have insisted in order to make the method clearer and more comprehensible to the student.

"Recent Advances in the Treatment of Mental Diseases" is the title of Chapter 21 by G. K. Yacorzynski. Shock treatment, prefrontal lobotomy (with a good discussion, of considerable interest, attempting to synthesize both from the standpoint of physiological theory), narcosynthesis, hypnotism, psychosomatic medicine, and experimental neurosis are covered briefly. It is surprising to read (on p. 542) the following: "One disadvantage of a psychologist serving as a psychotherapist is that he may lose sight of his primary objective, that of contributing to our knowledge and understanding of mental illness. . . . The practical beneficial results of therapy in no way indicates that the therapy taps one of the major processes of human adjustment. Theories of human behavior built on such evidence are to be accepted with great caution." However, the psychologist who acts as a psychotherapist may make some penetrating observations on behavior and thus contribute to understanding, i.e., research. Orientation towards the practical aims of therapy, for example, provided us with a very influential theory and research stimulant, viz., psychoanalysis. One may also question whether any benign therapy takes place *without* tapping a fundamental adjustive process.

Part V on "Professional Relationships" constitutes but one chapter (22), that concerning "Psychiatry and Clinical Psychology," by D. G. Wright, a psychiatrist. It considers the role of the psychologist, cooperating with the psychiatrist. Wright's discussion is eminently fair in its recognition of the psychologist's duties. In this part of the book, though, one would have liked to have seen a discussion of other professional relationships of the psychologist—in relation to the social worker, the educator and the industrialist, etc.

In summary, then, we have a book in which style of writing is variable, which in many places may be repetitious of work previously covered by the student in other courses, in which the treatment of the subject matter ranges from the very good to the mediocre, in which pertinent illustrative case material may be present or absent, but above all, in which the over-all editing appears to be somewhat lacking.

RALPH D. NORMAN.

University of Michigan.

BOOKS AND MATERIALS RECEIVED

ALMY, MILLIE CORINNE. *Children's experiences prior to first grade and success in beginning reading.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1949. Pp. viii+124. $2.35.

ANOMALY. *The invert.* (2nd Ed.) Baltimore: Williams & Wilkins, 1948. Pp. xxvi+289.

BRANHAM, VERNON C., AND KUTASH, SAMUEL B. (Eds.). *Encyclopedia of criminology.* New York: Philosophical Library, 1949. Pp. xxvii+527. $12.00.

BERLINER ANNA., *Lectures on visual psychology.* Chicago: Professional Press, 1948. Pp. iv+84.

BERGLER, EDMUND. *The basic neurosis.* New York: Grune & Statton, 1949. Pp. xii+353. $5.00.

DE BEAUVOIR, SIMONE. *The ethics of ambiguity.* New York: Philosophical Library, 1948. Pp. 163. $3.00.

DERI, SUSAN. *Introduction to the Szondi test.* New York: Grune & Stratton, 1949. Pp. xiv+354. $5.00.

DEUTSCH, FELIX. *Applied psychoanalysis.* New York: Grune & Stratton, 1949. Pp. 244. $3.75.

FLEMING, D. M. *Adolescence.* New York: International Universities Press, 1949. Pp. vii+262. $4.50.

GUTHRIE, EDWIN R., AND EDWARDS, ALLEN L. *Psychology: A first course in human behavior.* New York: Harper, 1949. Pp. x+315. $3.00.

HALL, VICTOR E., *et al. Annual review of physiology.* Stanford Univ., California: Annual Reviews, Inc., 1949. Pp. ix+643. $6.00.

HOLLINGSHEAD, A. B. *Elmtown's youth.* New York: John Wiley, 1949. Pp. xi+480. $5.00.

HURLOCK, ELIZABETH B. *Adolescent development.* New York: McGraw-Hill, 1949. Pp. x+551. $4.50.

JOAD, C. E. M. *Decadence.* New York: Philosophical Library, 1949. Pp. 430. $4.75.

KORNER, ANNELIESE F. *Some aspects of hostility in young children.* New York: Grune & Stratton, 1949. Pp. ix+194. $3.50.

LEWIN, KURT. *Resolving social conflicts.* New York: Harper & Bros., 1948. Pp. xviii+230. $3.50.

LORAND, SANDOR. *The yearbook of psychoanalysis.* Vol. 4. New York: International Universities Press, 1948. Pp. 356. $7.50.

MCMANUS, R. LOUISE. *The effect of experience on nursing achievement.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1949. Pp. viii+64. $2.10.

MELTON, ARTHUR W. *Apparatus tests.* Aviation Psychology Program, Research Reports. Report No. 4. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. xxi+1056. $3.50.

Plessner, H. *Über den schmerz.* Bern: Verlag Hans Huber, 1948. Pp. 182. Distributed by Grune & Stratton.

Rosenzweig, Saul (with the collaboration of Kate L. Kogan). *Psychodiagnosis.* New York: Grune & Stratton, 1949. Pp. xxi+380. $5.00.

Schafer, Roy. *The clinical application of psychological tests.* New York: International Universities Press, 1949. Pp. 346. $6.75.

Sellars, Roy W., McGill, V. J., and Farber, Marvin. *Philosophy for the future.* New York: Macmillan, 1949. Pp. xiii+657. $7.50.

Stouffer, Samuel A., *et al. The American soldier.* Vol. 1. *Adjustment during army life.* Princeton: Princeton Univ. Press, 1949. Pp. xii+599. $7.50.

Stouffer, Samuel A., *et al. The American soldier.* Vol. 2. *Combat and its aftermath.* Princeton: Princeton Univ. Press, 1949. Pp. 675. $7.50.

Suziki, Daisetz Tietaro. *Introduction to Zen Buddhism.* New York: Philosophical Library, 1949. Pp. 136. $3.75.

Swenson, Esther J., *et al. Learning theory in school situations.* Minneapolis: University of Minnesota Press, 1949. Pp. 103.

Snygg, Donald, and Combs, Arthur W. *Individual behavior.* New York: Harper, 1949. Pp. ix+386. $3.50.

Thompson, Charles E. *Thematic apperception test.* Cambridge: Harvard Univ. Press, 1949. $5.00 (Manual for test, $.50).

Senn, M. J. E. (Ed.) *Problems of early infancy.* New York: Josiah Macy, Jr. Foundation, 1948. Pp. 119.

Thorndike, Robert L. *Personnel selection test and measurement techniques.* New York: Wiley & Sons, 1949. Pp. viii+358. $4.00.

Van Der Lugt, M. J. A. *Adult psychomotor test series for the measurement of manual ability.* New York: Author, New York University, 1948. (Mimeographed.)

Van Der Lugt, M. J. A. *Psychomotor test series for children for the measurement of manual ability.* New York: Author, New York University, 1949. (Mimeographed.)

Van Der Lugt. M. J. A. *Un profil psychomotor.* Paris: Aubier, 1939. Pp. 123.

Wahl, Jean. *A short history of existentialism.* New York: Philosophical Library, 1949. Pp. 58. $2.75.

Waldfogel, Samuel. *The frequency and affective character of childhood memories.* Psychological Monographs, No. 291. Washington, D. C.: American Psychological Assn., 1948. Pp. 38.

Wertham, Frederic. *The show of violence.* New York: Doubleday & Co., 1949. Pp. 279. $3.00.

WITTY, PAUL. *Reading in modern education*. Boston: D. C. Heath, 1949. Pp. xvi+319.

ZIPF, GEORGE KINGSLEY. *Principle of least effort*. Cambridge, Mass: Addison Wesley Press, 1949. Pp. xi+573. $6.50.

L'Annee Psychologique. Edited by H. Pieron, A. Fessard and P. Fraisse. 45th and 46th Years (1944–1945). Paris: Presses Universitaries de France.

Guidance handbook for elementary schools. Division of Research and Guidance. Los Angeles: California Test Bureau, 1948. Pp. iv+158.

Helping students find employment. Washington, D. C.: Student Personnel Work, 1949. American Council on Education Studies. Series VI, No. 12, Vol. XIII.

Research frontiers in human relations. Philadelphia: American Philosophical Society, 1948. Pp. 325–410. $1.00.

Understanding the nervous veteran. Minneapolis: In-Service Training Section, Personnel Division, V. A. Regional Office, 1948. Pp. 21.

BOOKS REVIEWED IN THIS ISSUE

ALSCHULER AND HATTWICK'S *Painting and personality*, reviewed by L. J. STONE, 327.

BLUM'S *Industrial psychology*, reviewed by W. A. KERR, 311.

BRAY'S *Psychology and military proficiency*, reviewed by GLEN FINCH, 318.

EWEN'S *Mental health*, reviewed by L. IMMERGLUCK, 323.

EYSENCK'S *Dimensions of personality*, reviewed by LEONA TYLER, 319.

GESELL AND AMATRUDA'S *Developmental diagnosis*, reviewed by MARGARET KUENNE, 325.

HOLLINGWORTH'S *Psychology and ethics*, reviewed by G. W. HARTMANN, 313.

MACKINNON AND HENLE'S *Experimental studies in psychodynamics*, reviewed by H. H. KENDLER, 322.

MENDEL'S *Personality in handwriting*, reviewed by W. WOLFF, 324.

PENNINGTON AND BERG'S *Introduction to clinical psychology*, reviewed by R. D. NORMAN, 330.

PETERSON, MARZOLF AND BAYLEY'S *Educational psychology*, reviewed by G. M. BLAIR, 317.

SECHENOV'S *Selected works*, reviewed by E. G. BORING, 309.

TYLER'S *Psychology of human differences*, reviewed by R. L. THORNDIKE, 326.

VAUGHAN'S *Social psychology*, reviewed by R. A. LITTMAN, 314.

WHITE'S *The abnormal personality, a textbook*, reviewed by T. W. RICHARDS, 321.

Vol. 46, No. 5 September, 1949

Psychological Bulletin

STIMULUS GENERALIZATION OF CONDITIONED RESPONSES[1]

GREGORY RAZRAN
Queens College

INTRODUCTION

Historical. The doctrine of CR generalization and its presumed neurophysiological mechanism of irradiation and concentration entered Pavlov's system of behavior or, in his own words, "highest nervous activity," rather late. Extinction, the four kinds of internal inhibition (extinctive, differential, delayed, and conditioned), external inhibition (later discarded), inhibition of inhibition, and even analyzers, differential conditioning by contrasts, conditioning of higher order, and compound conditioning—all preceded it as concepts or phenomena. However, when irradiation-concentration was promulgated (**34**, 1910), it assumed a paramount position. Forming the core of all cortical dynamics, it became the "basic law of the highest[2] nervous activity" (**34**, p. 245) subsuming all older concepts and adding new ones, as subsidiaries, to account for alleged new CR discoveries. As is well known, Pavlov's entire system is conceptualized primarily upon terms borrowed, nominally, from classical neurophysiology. But one may also discern in his constructs the influence of theories of hydraulics and of sound transmission, the psychophysiology of G. H. Lewes, the logic and mental chemistry of J. S. Mill, and the stream of consciousness of William James.[3]

---

[1] This study was prepared for publication while the writer was a Fellow of the John Simon Guggenheim Memorial Foundation.

[2] For convenience, the writer's references are to the English translations of the works of Pavlov and Bekhterev. The writer himself, however, uses the Russian originals, and his translations may thus differ in some way from those published. In this particular case, the Russian word VYSSHEY—a word that occurs also in the subtitle of the Pavlov book and quite commonly in the text—is translated by Gantt as *higher*. The writer used *highest* in abstracting the book later translated by Gantt, and then found that the German translators used *höchste*.

Pavlov's empirical findings of CR generalization in dogs were supposedly confirmed by Bekhterev, the neurologist and psychiatrist, in human adults (4), and by Krasnogorski, the pediatrician, in children (26). And both Bekhterev and Krasnogorski fully accepted Pavlov's neurophysiological interpretations (*op. cit.*). However, Beritoff, the Georgian (U.S.S.R.) physiologist, and Konorski, the Polish biologist, were highly critical of Pavlov's cortical constructs (5, 25), and it is known to the writer that a number of Russian psychologists were dubious even about some of his findings. Still, Pavlov's prestige was very high, and opposition to his views—except by Beritoff—was very much subvocal in nature. Criticisms came later, in the Thirties, but then they were not on experimental grounds. The teachings of Pavlov and Bekhterev were declared to be not accordant with Marxism: materialistic, yet asocio-political and insufficiently dialectical. The criticisms were directed mostly against the Human Reflexology of Bekhterev—even though Bekhterev himself wrote a long monograph (3) attempting to prove that Reflexology is Marxian in essence. Pavlov's laboratory, experimenting with animals, was given greater subsidies. Marxists, like Cartesians, draw sharp dichotomies between human beings and animals.

In this country, in the early days of Behaviorism, CR generalization was taken as a qualitative fact, and even demonstrated by Watson—who called it transfer—in his emotional conditioning of Albert (59). But when Pavlov's books came out in English (33, 1927; 34, 1928), a number of American students of behavior became critical of Pavlov's theories and skeptical of a good many of his facts. Lashley, himself a pioneer in CR experimentation, led in general criticism (28); Loucks, a young and able CR experimenter, took to task the doctrine of irradiation and indirectly that of generalization (30, 31); and Guthrie, an old-time Behaviorist, offered a new theory of generalization and of conditioning as such (11, 12, 13). On the other hand, Hull, quite affected by the quantitative and systematic impliations of Pavlov's books, set up a special experiment (with M. J. Bass) (2) to test Loucks' scathing analysis of CR irradiation and generalization. Hull was apparently satisfied that his results redeemed Pavlov's findings and accorded CR generalization systematic status as *postulate 5* in his system (**20**). Generalization, in general, became fashionable at Yale, in the mid-Thirties. Hovland started a series of experiments in the field (**16, 17, 18**); and Spence embarked upon a scheme of "deducing" discriminative learning from CR generalization (**55, 56, 57**). Hilgard and Marquis allotted an approving half-chapter to generalization in their book (**15**), and other textbooks followed in accepting it as an established fact, presumably an objective and quantitative substitute for the old "Law of Similarity" advocated by J. S. Mill and by Spencer but disputed by most other Associationists (**8, 54, 58**).

---

[3] From a few long conversations with Pavlov in the summer of 1934, the writer gained the impression that Pavlov was quite conversant with the writings of Wundt, James, and the British Associationists (Pavlov studied at, but did not graduate from, the Ryazan Theological Seminary). However, despite his gallant tribute to Thorndike (spelled Thorndyke in Anrep's translation— **33**, p. 6) as his predecessor in the objective analysis of behavior and his discussion of the views of Guthrie, Lashley, and Köhler in one of his articles (**35**), it is the writer's opinion that Pavlov had little familiarity with modern American comparative and systematic psychology.

The Lashley-Hull controversy. The entire problem of CR generalization sprang up, however, again two years ago with a very stimulating article by Lashley and Wade (**29**). Lashley and Wade challenged everything in the doctrine: its facts, its interpretations, its significance. They cited Loucks' analysis, apparently not only for its specific criticisms, but also for its general implications that conclusions drawn by Pavlov and his students ought not to be taken at their face value. They stated that Bass and Hull (**2**) and Hovland (**16, 17, 18**) did not really corroborate primary stimulus generalization inasmuch as they used human subjects "for whom the stimulus series represented familiar relational sequences" and who may have used "habits of relational thinking" (p. 75). They mentioned the study by Wickens (**60**) who "failed to demonstrate a gradient of stimulus generalization" and one by the writer (**48**) who found "generalization with different types of stimuli too variable to formulate under any simple laws." Finally, Lashley and Wade detailed a series of their own experiments in all of which they failed to find generalization in animals or human subjects by their own technique of "training a group of subjects in a reaction to a single stimulus then opposing that stimulus to another on the same stimulus dimension and comparing the rates of formation of a discriminative habit when the reaction to the initial stimulus is reinforced and when it is extinguished by the differential training."

Lashley and Wade's interpretations are even more far-reaching than their criticisms of the experimental findings of CR generalization. These interpretations may be best summarized in their own words. They are:

> 1. There is no "irradiation" or spread of effects during primary conditioning (p. 74).
>
> 2. The phenomena of "stimulus generalization" represent a failure of association (p. 74) . . . failure to note the distinguishing characteristics of the stimulus or to associate them with the conditioned reaction (p. 81), . . . [and this generalization is thus really a] generalization by "default" (p. 82).

3. The "gradient of habit strength" is a product of variable stimulus thresholds, not a spread of associative process (p. 74). . . . The gradient varies with degree of attention and is unrelated to habit strength. . . . When inattention or threshold values are not involved, no evidence of a gradient is found. . . . Consequently, a test for "irradiation" may give the appearance of a gradient of habit strength when it is actually measuring discriminative thresholds under distraction (p. 84).

4. It [true generalization] does not occur in conditioning to a single stimulus but is somehow a function of differential training with two or more stimuli on the same dimension (p. 82). . . . The "dimension" of a stimulus series . . . do not exist for the organism until established by differential training (p. 74). . . . The dimension itself is created by or is a function of the organism and only secondarily, if at all, a property of the physically definable character of the stimuli (p. 82). . . . In the early stages of Pavlovian conditioning the only "dimension" common to such [tested for conditioning] stimuli is that all produce a sudden change in the environment. . . . With continued training the subject . . . may or may not show narrowing of the effective range on a stimulus dimension. Apparently such changes are a matter of chance noting of differences, generally with little regularity [reference to a study by the writer (**48**)] (p. 81).

Hull replied to Lashley and Wade (**21**). His reply is much more a defense of the experimental findings of CR generalization than a discussion of Lashley and Wade's special interpretations. He cites a private communication by Wickens that his [Wickens] data were in some respects quite harmonious with Hovland's findings and only when the Chi Square was calculated did "none of these differences [the differences between the conditioned and generalization stimuli] reach the 10 per cent level" (p. 130). He further states that in some of Hovland's experiments "the gradient at the first trial was horizontal" and that "If the gradient eventually found were due to the indirect effect of previously formed habits, e.g., of speech, it should have appeared at the very first trial" (p. 128). Hull then divulges unpublished results by Spence who found a stimulus generalization gradient with a Lashley and Wade technique, and reproduces seven graphs of CR generalizations from the studies of Anrep (**1**), Bass and Hull (**2**), Hovland (**16, 17**), Brown (**7**), and Wickens (**60**).[4] Hull's own views need not be gone into here since they are well known from his writings on this topic (19, 20, 21, 22). Hilgard (14, 1948) seems now to lean more to Lashley and Wade's views than to those of Pavlov and Hull in stating that "generalization may be along discriminated dimension" (p. 338). But a nearly straight line generalization curve with a modified Lashley and Wade technique was found by Grandine and Harlow (9) who, however, state that their findings "give no definitive answer as to whether this phenomenon of stimulus CR generalization is the result of some preestablished gradient around the training point, or is the specific resultant of the habit or habits established" (p. 336).

---

[4] Hull states: "The reference given by Lashley and Wade to Razran has been gone over with great care, but the present writer has been quite unable to find anything which would cast doubt on the reality of the falling gradient of stimulus generalization. . . . The article is essentially a contribution to a controversy concerned with response to explicit patterns of stimulations, a matter which is not under consideration here" (**21**, p. 130). Hull is obviously discussing another study by the writer (**43**) which is given in Lashley and Wade's bibliography, and not the particular one (**48**) to which they refer in the text. There is a typographical error in Lashley and Wade's reference: the writer's

Background of the writer's interest. The writer has been interested in CR generalization since he published (with C. J. Warden) his first survey of pertinent Russian literature in 1929 (52). Through that survey and a number of others, the writer became not only critical of the theories of Pavlov and Bekhterev (39–42) but also convinced that the very empirical findings of the Russian laboratories need considerable checking and analysis. He came to learn that while the laboratories of Pavlov and Bekhterev are well advanced with respect to apparatus, physical set up of stimuli administration, controls of secondary cues, and refined measurements, their workers are very naive with regard to experimental design, sampling errors, and statistical treatments in general. Furthermore, it became known to him that individual CR experimenters in Pavlov's laboratory seldom exercise independence in the interpretation of their own data. Consequently, when the writer wished 10 years ago to review critically some significant aspect of conditioning, he chose extinction (45) rather than generalization, believing a good deal of the alleged phenomena of the latter to be little established. In 1937–1940 he carried out a series of experiments on the generalization of salivary CR's in adult human subjects. Three of the experiments have been published (44, 47, 48), but the publication of the rest has been delayed (three are in press now). In all three published studies, some CR generalization was manifested, but its course and very existence varied so much with the type of stimulus used that the writer stated in 1940 that his "empirical findings cast grave doubt upon attempts to 'deduce' transposition in discrimination experiments from generalization" and that laws of transfer "must be patiently discovered through experimentation and study of results at each level or form of stimulus, response, and organism organization" (48, p. 11).

---

mentioned study should have been referred to as "(19)" rather than "(18)." This study does not appear among the references in Hull's article.

Purpose of the present review. While the alleged empirical evidence for irradiation—the temporal after-effects of the application of conditioned stimuli—is, in the writer's opinion, worth checking and cannot be considered settled by Loucks' analysis,[5] and while response generalization offers interesting possibilities,—there is no doubt that the important problem for students of behavior is that of stimulus generalization, or the more or less permanent CR influence of a conditioned stimulus upon a non-conditioned stimulus that is in some way related to it. And this is the problem which the present article purports to treat as fully as feasible. Specifically, the article will attempt, first, to examine all available evidence in answer to four empirical questions (three primary and one secondary) on CR generalization, and, then, to provide some integrative interpretations of the evidence and the answers.

The three primary empirical questions are:

1. Is there consistent evidence to show that an organism conditioned to some stimulus will also produce a CR in some degree to some other related stimulus, with which the organism has had no previous pertinent experience?
2. If the answer to the first question is positive, What is the relation of the strength of the CR to the related stimulus—magnitude, latency, resistance to extinction—to the strength of the CR to the conditioned stimulus?
3. Is there a gradient of CR generalization? And if there is, what does it correlate with? Is the correlation of such a nature as to permit some tentative mathematical treatment of the gradient?

The secondary empirical question is:

What is the relation between the amount of training that a CR has received and the answers to the first three primary questions? (Another secondary empirical question, the generalization of extinction cannot, unfortunately, be given consideration here, for lack of space.)

The evidence to be examined. The evidence for answering the empirical questions and for subsequent integrative interpretations of the answers will be sought primarily in (a) the studies from Pavlov's laboratory of salivary CRs in dogs (67 studies), (b) Yale studies of conditioning the GSR in adult human subjects (five studies), and (c) the writer's studies of salivary CRs in adult human subjects. Salivary conditioning of dogs constitutes probably more than 90 per cent of the CR work in the U.S.S.R., while there is reason for excluding from the present comparative analysis four other American studies, with data on

---

[5] A preliminary analysis by the writer of a number of Russian studies, together with a closer re-examination—and some re-analysis—of Loucks' analysis, indicate that these after-effects certainly occur more often that chance would warrant. However, the writer agrees fully with Loucks that they are not of a nature to prove irradiation. They probably are results of some very initial differential conditioning.

CR experimentation. Of the four, the study by Wickens (60) is complicated by involving response as well as stimulus generalization and by the general vicissitudes of conditioning finger withdrawal. The experiment by Brown (7), besides being instrumental rather than classical conditioning, used intensities as generalization stimuli without being able to control the factor of intensity *per se* (see later sections of the present article). The two studies by the Lashley and Wade technique (9, 29) will be utilized but might be better kept apart with respect to direct comparisons. On the other hand, the Pavlov, the Yale, and the writer's studies have all used autonomic responses and classical techniques and have, in addition, the advantage that in the Yale studies the stimuli and in the writer's studies the response were the same as in Pavlov's studies, the major source of evidence.

THE EVIDENCE FROM PAVLOV'S LABORATORY

Source material. Prior to 1924, most—but not all—full reports of CR work in Pavlov's laboratory were published in the form of doctoral theses, in fulfillment of the requirements for the M.D. degree at the St. Petersburg Military Medical Academy where Pavlov served first as Professor of Pharmacology and then as Professor of Physiology. These theses are usually 100 to 200 pages long and as a rule represent, in the writer's estimate, about as much work as Ph.D. theses in physiology or psychology in a first-rate American university. Their tables, or rather protocols, are very detailed, trial-by-trial presentations of magnitudes and latencies of salivation, and of amounts and qualities of food seeking motor accompaniments, for each stimulus and each dog. The writer knows of 49 such CR theses from Pavlov's laboratory and he has read most of them (two of the six reports in Loucks' analysis are such theses).

Since 1924, however, the bulk of CR experimentation from Pavlov's laboratories has been published in the *Trudy Fiziologicheskikh Laboratorii Akademika I.P. Pavlova* (Transactions of the physiological laboratories of Academician I.P. Pavlov). The writer has in his possession ten volumes of these *Trudy*, from 1924 to 1941. These ten volumes contain 245 separate CR studies (nearly all salivation in dogs). The studies are as a rule less extensive than those in the earlier doctoral theses and, moreover, are not presented in great detail (the total number of pages in the ten volumes is a little over 3000). They were performed, not by candidates for the M.D. degree, but by Pavlov's numerous research assistants and associates, often physiologists of note in their own right.

Originally, the writer intended to analyze first the material of CR generalization in the doctoral theses. But, unfortunately, these theses are not available to him this year, and he decided to confine himself to the data in the ten volumes of the *Trudy*. However, he does have impressions of the theses which he read prior to the preparation of this article.

Method of analysis. Only very few of the 245 CR studies in the *Trudy* deal specifically with CR generalization—or irradiation—which Pavlovians consider to have become established facts years ago. But a good number of the studies contain exact CR generalization data, incidental to the solution of some other problem in "cortical dynamics." Very often a CR experimenter, after he has formed a CR to some stimulus, will try out the CR with some other non-conditioned stimulus or stimuli, and these other stimuli are quite commonly related in some ascertainable way to each other and to the conditioned stimulus. The writer has thus gone over a few thousand tables in the *Trudy*, and every time that he noted that a CR was tried with a non-conditioned stimulus, he computed the magnitude of its salivation as a per cent of the salivation to the conditioned stimulus, for the particular session and particular animal. In computing the percentages, account was of course taken of the equality of other experimental factors in the session; and to avoid the problem of differential conditioning, only the first two trials of a CR to a non-conditioned stimulus were used. Moreover, inasmuch as most experimenters from Pavlov's laboratory present in each experiment the CR history of each dog, generalization stimuli that appeared in such histories were excluded from the computations.

As the data began accumulating, the writer decided to limit his task to four kinds of *dimensional* (other than intensity) CR generalizations, three kinds of *intensity* generalizations, and a number of what, for lack of another name, may be called *inter-dimensional* and *inter-sensory* CR generalizations. The four kinds of dimensional (other than intensity) generalizations included (a) frequencies of beats of metronomes, (b) frequencies of tones, (c) frequencies of rhythmic tactions, and (d) spatial distances of tactions. The three kinds of intensity generalizations contained intensities of flashes of lights, intensities of bells, and intensities of whistles. The inter-dimensional and inter-sensory CR generalizations refer to generalizations between such stimuli as metronomes and whistles, metronomes and lights or tactile stimuli, and the like.

More specifically, data were obtained for the following conditions:

1. All the 12 permutations of generalizations between metronomes, whistles, light flashes, and tactile stimuli;
2. Four dimensional (other than intensity) *positions*, with regard to fre-

quencies of metronomes, tones, and rhythmic tactions, and to spatial distances of tactions;

3. Three lower and three higher intensity positions, with respect to lights, whistles, and bells.

The dimensional and intensity positions were determined, with two exceptions, from the objectively given stimuli characteristics: frequencies, distances, c.p.s.'s, and decibels. The two exceptions were some cases in which spatial distances had to be estimated by the writer, and some cases in which the intensities of bells and whistles were not given in decibels (only the experimenters' mere descriptions of "weaker," "still weaker," "weakest," "stronger," "still stronger," and "strongest" being included). On the other hand, it should be pointed out that the magnitude of the distances between positions within dimensions and intensities were as a rule unequal, and that averaging had to be done by combining corresponding positions (i.e., I with I, II with II, III with III, and IV with IV).

The writer is fully aware of some of the pitfalls of his analysis. He would like to point out, however, first, that the Pavlov laboratories are extremely routinized with regard to experimenter, design, apparatus, stimuli, and responses; and, second, that extreme care was taken by him to see that each of the nearly 700 comparisons between the CR's to the generalization and to the conditioned stimuli was performed under wholly equal conditions. Moreover, while a combined analysis may have masked the effects of some variables, one of these variables, the amount of training that the CR has received, has been singled out for special study. The data from frequencies of metronomes, spatial distances of tactions, intensities of lights, and transfers from metronomes to lights have been fractionated so as to compare CR generalization after 1–20, 21–40, 41–100, 101–300, and more than 300 reinforcements. Similar treatments of other variables—which the Russians maintain affect CR generalization, such as the age of the animals, cortical extirpation, certain drugs, and a few more—would have been possible but were not considered worth while undertaking, at least for the time being. It is believed, though, that the alleged effects of these variables might well have been cancelled out, since the writer included in his combined analysis data from both supposed generalization-increasing and supposed generalization-decreasing studies.

Results for Dimensional (other than intensity) CR Generalization

All the *Trudy* data on dimensional (other than intensity) CR generalization are combined in Table I. Each entry in the table is a mean per cent of generalization of a number of determinations (exact number is given in parentheses), obtained from 8 to 67 different experiments. The *t*'s and *P*'s of the entries in Table I—as well as of the entries in the subsequent tables—have all been ascertained, but their separate tabulation would unduly lengthen the article. The most pertinent ones will be taken up in the text, in connection with each finding.

TABLE I

DIMENSIONAL (OTHER THAN INTENSITY) STIMULUS GENERALIZATION OF SALIVARY CONDITIONING IN DOGS. DATA FROM 67 DIFFERENT EXPERIMENTS IN PAVLOV'S LABORATORY

Each entry is a mean per cent of conditioned salivation to the non-conditioned generalization stimuli. Figures in parentheses are numbers of determinations.

Conditioned Stimuli	*Steps Removed from Conditioned Stimuli*			
	I	II	III	IV
Metronome	61(28)	57(22)	43(16)	45(11)
Tone	69(23)	59(19)	46(13)	42(10)
Spatial Taction	78(18)	59(14)	63(9)	54(8)
Rhythmic Taction	51(16)	57(13)	41(8)	44(8)

As seen from Table I, the answers to the first two of the previously posed "primary empirical questions" (*supra*, p. 342) are unmistakable. Non-conditioned generalization stimuli do evoke conditioned responses, and these responses are smaller in magnitude than those evoked by the conditioned stimuli. All the *t*'s between the magnitudes of the conditioned and the generalization CR's are significant with *P* equalling .01, the smallest *t* being 17.4. However, the answer to the third question, the question of a CR gradient, is by no means certain. Four of the 12 differences between adjacent generalization stimuli are reversals, and four of the remaining eight differences are insignificant, with *P* equalling .05. Yet *there would be no reversals and differences between adjacent generalization stimuli would be significant if the four dimensional steps were reduced to two by combining the first step with the second and the third with the fourth.* The importance of the last statement, as well as the fact that two of the four reversals occurred between the third and the fourth dimensional steps, will be discussed in a later section.

Results for Intensity CR Generalization

All the data on CR generalization of intensities are contained in Table II. They are given in percentages, and separately for each of the three lower and each of the three higher intensities. Unlike Table I, the gradients in Table II are quite consistent. There are no reversals, and all 12 differences between adjacent stimuli are significant with P equalling .05. However, Table II shows *two* gradients: a descending one for lower intensities but an ascending one for higher intensities. Indeed, it is very doubtful whether intensity generalizations could be at all considered true CR generalizations (cf. Hovland, **17**). At any rate, these generalizations are certainly *sui generis*, and little ought to be inferred from them to generalizations within non-intensity dimensions. Furthermore, there is reason to believe that an intensity-like factor, namely, psychological intensity resulting from summation of stimuli, masks true generalizations along such stimuli dimensions as frequencies of metronomes and of tactile vibrators.

TABLE II

STIMULUS INTENSITY GENERALIZATION OF SALIVARY CONDITIONING IN DOGS. DATA FROM 54 DIFFERENT EXPERIMENTS IN PAVLOV'S LABORATORY

Each entry is a mean per cent of conditioned salivation to the non-conditioned generalization stimuli. Figures in parentheses are numbers of determinations.

Conditioned Stimuli	*Lower Intensity Steps*			*Higher Intensity Steps*		
	I	II	III	I	II	III
Lights	79(14)	69(11)	58(8)	118(14)	128(9)	149(8)
Whistles	68(12)	58(11)	49(9)	137(13)	149(8)	165(8)
Bells	72(11)	64(9)	56(7)	124(10)	138(8)	149(6)

Results for Inter-Dimensional and Inter-Sensory CR Generalizations

What is striking about this type of CR generalization, as disclosed in Table III, is its large amount. There is, for instance, more CR generalization from a metronome to a whistle than from one metronome to another three steps removed, and there is nearly as much generalization from a light to a whistle than from one tone to another four steps away.[6] These results, among others, cast grave doubt, in the writer's opinion, upon any postulate that the magnitude of CR generalization merely varies with the degree of relatedness of the generalization stimuli to the conditioned stimuli in some stimulus dimension or dimensions. A better case could perhaps be made out for a variation with some denotative relatedness or relatednesses in a phenomenal world.

---

[6] It should be mentioned, however, that it is very much easier to establish differential conditioning between inter-dimensional and inter-sensory stimuli than between stimuli on the same dimension.

Results for CR Generalization as a Function of the Amount of Training of the CR

With the exception of intensity generalization which was already noted as a special case, all the data in Table IV point to the following conclusion: (a) that CR generalization increases in the very initial stages of training the CR; (b) with further training of the CR, it begins to decrease slowly; and (c) after a great number of reinforcements, the generalization may increase again. All the *t*'s between 1–20 and 21–40 reinforcements are significant (with *P* equalling .01), while the *t*'s between 21–40 and 41–100 and those between 101–300 and more than 300 reinforcements are significant (with *P* equalling .05), for frequencies of metronomes and for spatial distances.

TABLE III

INTER-DIMENSIONAL AND INTER-SENSORY STIMULUS GENERALIZATION OF SALIVARY CONDITIONING IN DOGS. DATA FROM 58 DIFFERENT EXPERIMENTS IN PAVLOV'S LABORATORY

Each entry is a mean per cent of conditioned salivation to the non-conditioned generalization stimuli. Figures in parentheses are numbers of determinations.

Conditioned Stimulus	*Generalization Stimulus*	*Per Cent of Generalization*
Metronome	Whistle	44(33)
Metronome	Light	38(31)
Metronome	Rhythmic taction	40(27)
Whistle	Metronome	42(21)
Whistle	Light	36(19)
Whistle	Rhythmic taction	38(20)
Light	Metronome	38(22)
Light	Whistle	39(24)
Light	Rhythmic taction	34(18)
Rhythmic taction	Metronome	41(19)
Rhythmic taction	Whistle	36(21)
Rhythmic taction	Light	32(17)

Summary of the Evidence from Pavlov's Laboratory

A statistical analysis of the data on CR generalization contained in 67 experiments, performed between 1924 and 1941, in Pavlov's laboratory provides full and clear answers to the first, second, and fourth questions posed in the first section of this article but offers only a partial

TABLE IV

STIMULUS GENERALIZATION OF SALIVARY CONDITIONING IN DOGS AS DEPENDENT UPON THE NUMBER OF REINFORCEMENTS OF THE CONDITIONED STIMULUS. DATA FROM 67 EXPERIMENTS IN PAVLOV'S LABORATORY

Each entry is a mean per cent of conditioned salivation to the non-condition generalization stimuli. Figures in parentheses are numbers of determinations.

Types of CR Generalizations	*Number of Reinforcements of the Conditioned Stimulus*				
	1–20	21–40	41–100	101–300	300 up
Frequencies of Metronomes	39(12)	64(17)	55(16)	49(12)	58(15)
Spatial Distances of Taction	54(8)	72(10)	64(9)	61(7)	74(11)
Metronomes to Lights	31(6)	42(8)	41(5)	36(5)	39(7)
Intensities of Lights*	96(10)	99(11)	103(14)	100(12)	96(17)

* Combining lower and higher intensities.

and not wholly certain answer to the third question. Specifically, th answers are:

1. Dogs conditioned to secrete saliva upon the application of some conditioned stimulus will also produce conditioned salivation upon the application of related stimuli with which the animals have had no previous pertinent experience.

2. Unless the intensity of the related stimulus is considerably higher than that of the conditioned stimulus, the magnitude of the related generalization CR will be considerably smaller than that of the trained CR.

3. CR generalization increases in the very initial stages of training the CR but upon further training begins to decrease slowly, while after a large number of reinforcements (over-training) it may increase again.

4. The gradient of CR generalization is very crude, consisting only of two or three steps.

5. The gradient does not vary *merely* with the degree of relatedness of the generalization stimuli to the conditioned stimuli along some stimulus dimension or dimensions.

THE YALE STUDIES

All the Yale studies to be analyzed here have been performed with human subjects and with the GSR as the conditioned response. Unlike the studies from Pavlov's laboratory, the Yale investigations have mastered, as might have been expected, not only the physical setups but also the design and the statistical treatments of their problems. Indeed, Hovland's three experiments may well be held up as CR Ph.D. paradigms. Moreover, there is little doubt that individual experimenters at Yale exercise much more independence in interpretations than individual experimenters in Pavlov's laboratory.

Hovland's Experiment with Frequencies of Tone (16)

This experiment has been quoted so often as an exemplar of a lawful and consistent CR gradient that the writer takes the liberty to reproduce here, for purposes of analysis, its main, and practically only, table (there are two more single line tables in the article). Hovland's table becomes Table V in this article, and the reader is asked to examine it very carefully.

TABLE V*

GENERALIZATION OF EXCITATORY TENDENCIES

Amplitudes of galvanic responses (in mm.) to conditioned tone (0) and to tones 25(1), 50(2), and 75(3) j.n.d.'s removed in frequency. Each value is average of two determinations following 16 reinforcements. Subjects I–X were conditioned to tone of 153 cycles; subjects XI–XX to tone of 1967 cycles.

Subject	*Tonol Stimuli*			
	0	1	2	3
I	16.2	11.3	12.6	11.4
II	22.4	18.1	25.4	20.9
III	13.5	6.7	11.2	6.3
IV	15.3	6.4	3.3	7.6
V	19.2	18.5	22.8	15.3
VI	16.2	18.5	11.9	13.9
VII	23.7	17.1	18.0	16.5
VIII	11.5	14.1	9.6	13.8
IX	13.8	10.3	13.4	9.9
X	22.4	18.7	10.7	15.3
XI	23.7	21.3	21.1	10.2
XII	16.5	20.6	13.9	12.4
XIII	17.7	18.4	15.2	13.7
XIV	18.6	13.9	14.3	12.5
XV	15.9	15.5	13.8	14.2
XVI	18.8	12.3	14.6	9.7
XVII	21.3	9.7	10.5	12.3
XVIII	23.2	17.8	13.9	14.5
XIX	13.8	16.8	9.3	11.8
XX	19.9	12.3	6.9	15.6
Mean	18.3	14.91	13.62	12.89
P.E.M.	0.57	0.64	0.8	0.49

* Reproduced from Hovland with permission.

As seen from Table V, the mean magnitude of the CR to the conditioned stimulus was 18.3 ± 0.57 (P.E.) mm., while the mean magnitudes of the generalization CR's—removed 25, 50, and 75 j.n.d.'s from the conditioned stimulus—were respectively: 14.91 ± 0.64, 13.6 ± 0.80, and

12.89±0.49 mm. The differences between the magnitude of the conditioned CR and the generalization CR's are fully reliable statistically, and so is the difference between the CR to the first and to the third generalization stimulus (Diff./P.E.$_{\text{diff}}$=4.21). *But the differences between adjacent generalization stimuli, i.e., between the first and the second and between the second and the third are not reliable statistically and are rather small.* Furthermore, if we examine the records of individual subjects, we find the following:

1. Nineteen of the 20 subjects showed one or more reversals, i.e., had mean CR's to more remote stimuli greater than to less remote stimuli;
2. Seven of the 19 subjects had mean CR's to generalization stimuli greater than the mean CR's to the conditioned stimulus;
3. The 20th subject, who showed no reversals, hardly manifested any gradient between the second and the third generalization stimuli—respective CR's being 21.3 and 21.1 mm.

Thus, not a single of Hovland's 20 subjects revealed a consistent gradient of CR generalization and only one of the three gradient-determining differences was statistically reliable.

Hovland's Study with Varying Intensities of Tones (17)

Two equated groups of 16 subjects each were used. The conditioned stimuli were a tone of 86 decibels for one group and a tone of 40 decibels for the other group, while the generalization stimuli were three descending tones 50, 100, and 150 j.n.d.'s removed from the conditioned stimulus for the first group, and three similarly removed ascending tones for the second group. Each conditioned stimulus and each generalization stimulus was tested three times, after 16 reinforcements of the conditioned stimulus with the unconditioned stimulus of electric shock. The mean CR to the conditioned tone of 86 decibels was 17.85 mm. while the respective CR's to the descending generalization tones were: 13.37, 10.94, and 8.9 mm. But the mean CR to the conditioned tone of 40 decibels was 10.3 mm. while the respective generalization CR's were: 12.98, 14.3, and 15.75 mm. Combining the results of the two groups, Hovland obtained a mean CR of 14.3±0.6 to the conditioned tones and means of 13.7±0.57, 13.17±0.57, and 12.62±0.44 to the three non-conditioned generalization tones. Hovland states that such combining "enables one to determine the gradient of *intensity* generalization with the intensity effect *per se* held constant" (p. 282) and believes that it is justifiable "because the relationship between intensity and magnitude of response conditioned is linear" (p. 285). This linearity cannot be considered wholly established. But even, if it were established, this gradient of intensity generalization with intensity *per se* held constant is certainly, as Hovland's data indicate, very small and very unreliable statistically.

Hovland's Study of CR Generalization as a Function of Amount of Reinforcement (18)

Hovland's findings here are so much like those from Pavlov's laboratory—namely, an increase in generalization in early stages of training the CR, then a slow decrease and then again an increase—that they need not be restated. The only difference between the types of studies is that the decrease in generalization began in Hovland's subjects after 16 reinforcements and in Pavlov's dogs after about 40 reinforcements, a difference that is only apparent, since more reinforcements are as a rule required to form a salivary CR in a dog than a conditioned GSR in a man.

Humphreys' Experiment (23)

Humphreys used two methods of reinforcement, a regular and a special, in his experiment with conditioning frequencies of tones. With the regular method, used for 34 subjects, the conditioned stimulus was always reinforced; while with the special method, used for 20 subjects, only half of the applications of the conditioned stimulus were accompanied by reinforcement (electric shock). One of Humphreys' three generalization stimuli was 5 j.n.d.'s removed from the conditioned stimulus (a tone of 1967 cycles), and another was 15 j.n.d.'s removed. The third generalization stimulus, on the other hand, was removed 25 j.n.d.'s from the conditioned stimulus in one half of the subjects. But in the other half it was a tone 26 j.n.d.'s removed, but the lower octave of the conditioned stimulus. Humphreys' results show the mean magnitude of the CR to the conditioned tone to have been 3.52 mm., with magnitudes of 2.91 for the tone 5 j.n.d.'s removed, 2.84 for the 15 j.n.d. tone, 2.45 for the 25 j.n.d., and 3.27 mm. for the lower octave —when the regular method of reinforcement was used. With the special method of 50 per cent reinforcement, the mean magnitude of the CR to the conditioned tone was 3.94 mm., while the mean magnitudes of the generalization CR's were respectively: 3.76, 4.19, and 3.13 mm. (The last figure is a combination of the CR's to the tone 25 j.n.d.'s removed and to the octave 26 j.n.d.'s away, no separate figures being given by the experimenter.) Thus, we have here an experiment on CR generalization that shows a reversal and very small and unreliable differences between CR's to generalization stimuli differing widely in dimensional relatedness. Hull (21) does not cite Humphreys' study in his marshalling of evidence for a true gradient of CR stimulus generalization. But Lashley and Wade (29) do mention it.

The Bass and Hull Experiment (2)

Bass and Hull worked with spatial generalization of taction, using eight subjects. The generalization stimuli were spaced approximately

16, 32, and 48 inches from the conditioned spot, and each generalization stimulus was tested 32 times. The mean CR to the conditioned stimulus was 5.74 ± 0.39 mm., while the means to the generalization stimuli were respectively: 5.63 ± 0.49, 4.75 ± 0.395, and 3.36 ± 0.32 mm. Four of the eight subjects showed reversals, and the subject for whom data are presented on four successive days showed reversals on two of the four days. In a certain way, however, the gradient obtained by Bass and Hull is the sharpest from the Yale laboratories. But, of course, it too—with reversals in half of the subjects and practically no difference between the group CR to the conditioned stimulus and to the generalization stimulus 16 inches away—could hardly support the conclusion by Hull that "This generalization gradient of reaction strength is a monotonic decreasing function of the magnitude of the differences between the conditiond and the unconditioned[7] [generalization] stimuli" (**21**, p. 133). Furthermore, there is a serious criticism of the Bass and Hull study, as far as the gradient is concerned, namely: their alternate reinforcement of the conditioned stimulus and non-reinforcement of the generalization stimuli permitted differential conditioning by contrasts to develop.

Summary and Evaluation of the Yale Studies

There is on the whole a striking similarity between the results on CR generalization in the Yale and in the Pavlov studies. Both provide about the same answers to the four questions asked in the first section of this article. The fact that the Yale studies were performed with human subjects makes their gradient results subject to interpretations of "habits of relational thinking" (Lashley, **29**) and of earlier experience with "conventional dimensions of discriminations" (Hilgard, **14**). Hull's statement about the "relatively non-voluntary character of galvanic skin reactions" (**21**) falls short in that in generalization it is the consciousness of the stimuli, not only of the responses, that counts, and human subjects are certainly conscious of the stimuli in GSR experiments, and consciousness of stimuli certainly affects the GSR.

However, while human subjects could use the devices mentioned by Lashley and by Hilgard, they of course need not use them, and the very crude and irregular gradients obtained could be construed as evidence that they did not, at least consistently, use them. But this may be answered by the argument that the "sets" in the Yale studies did not call for conscious and active discriminations, so that these have manifested themselves only occasionally and incidentally (something like "incidental memory" or "premature reactions" in reaction time experiments). On the other hand, the close similarity between the Yale and the Pavlov results should lend support to the supposition that previous experience was not much operative in the Yale studies. Ultimately the problem seems to resolve itself into the alternative of the Yale findings being results of (a) relational thinking superimposed upon a true CR gradient or (b) relational thinking superimposed upon chance, with the effects of relational thinking varying possibly from zero to 100 per cent.

---

[7] *Unconditioned* is a very confusing adjective here. *Non-conditioned* would be better.

RAZRAN'S EXPERIMENTS

All experiments on CR generalization by the writer were with salivary CR's in college undergraduates. His general technique of salivary conditioning has been described previously (**42**, **44**, **46**) and need not be repeated here. All that might be added are three of its more recent characteristics; viz.: (a) multiple intermittent one-second presentations of stimuli-to-be-conditioned during single continuous eating periods of two to four minutes, so as to provide maximum attention for the stimuli and to make the entire task more "molar" and meaningful; (b) misinforming the subjects about the nature of the experiment, so as to forestall disrupting subjective attitudes; and (c) varying the food and scheduling the sessions in late mornings and afternoons, so as to insure adequate psychophysiological motivation. Of the three characteristics, the most important one is, in a way, "misinforming the subjects" such as telling them that the experiment aims to study "the effects of hunger or satiety or digestion upon eye-fatigue" (with visual stimuli) or "upon ear-fatigue" (with auditory stimuli) or "upon memory" (with verbal stimuli), and actually administering some "sham" tests of fatigue or memory. Thus, the writer's subjects are very attentive to the stimuli, the responses, their own physiological and mental states; but wholly unaware of the experimenter's objective and his attempt to condition them. Their "consciousness" and "sets" have been "naturally" diverted. (Occasionally a subject will "catch on," and then his results are treated separately.)

Experiments with Sensory Stimuli

Two types of experiments were performed. In one experiment, used with 32 subjects, the conditioned stimuli were the flashes of two to ten miniature spherical lights ($\frac{3}{8}$ inches in diameter, 2 c.p., 2.5 volts) of the same or of different colors. The lights were arranged in different spatial patterns—many of the patterns being the same as the dot arrangements of Schumann, Rubin, and Wertheimer—and they were flashed either simultaneously or in specially selected temporal sequences. Their generalization stimuli were (a) fewer lights, (b) lights differing in color, (c) lights in different spatial arrangements, and (d) lights in different temporal sequences. In the other experiment, used with four subjects, the conditioned stimuli were 12 musical intervals (excluding the natural minor seventh) in the c^2–c^3 octave of a Stoelting *just* harmonium. The generalization stimuli were different musical intervals in the same octave and with one common tone, different musical intervals in the same octave but with no common tone, the same musical intervals in a different octave and in a different key, and the same musical intervals in a different octave but in the same key.

There was no attempt in these experiments to plot CR gradients but to a) establish the existence of CR generalization, in general; b) study its dependence upon amount of reinforcement of the CR; and particularly c) study the dependence of CR generalization upon (1) the spatial, temporal, and color patterns in the case of lights, and upon (2) common ratios, partials, and absolute tones in the case of the musical intervals. The results showed: (a) unmistakable evidence for the existence of CR generalization and for the lesser magnitudes of generalization CR's, (b) impossibility of predicting the course of CR generalization to the light patterns from either the "principles of patternization" of Wertheimer or the simple CR generalization of Pavlov, (c) the determining roles of ratios in the CR transfer from musical intervals.

Phonetographic Generalization (47, 49)

Four subjects were used. The conditioned stimuli were single English words, and the generalization stimuli were words related to the conditioned words in sound and spelling, a relatedness named by the writer phonetographic. A crude CR gradient was found here. Thus, the mean generalization to homophones (*urn*—"earn"; *style*—"stile"; *surf*—"serf"; *freeze*—"frieze") was 37 per cent, while the generalization from *flower* to "glower" was 35.1 per cent, from *dark to* "mark," 31.6 per cent, from *mock* to "dock," 27.2 per cent, from *flower* to "shower," 20.2 per cent, and from *day* to "may," 19.6 per cent. The *t*'s here are significant with *P* equalling .05, if the differences are more than four per cent.

Semantic Generalization (49)

Four to 11 subjects were used, and the conditioned stimuli were single English words. The generalization stimuli were word related to the conditioned words semantically; synonyms, contrasts, coordinates, supraordinates, subordinates, whole-part's, part-whole's, and so forth. The mean generalization to the synonyms was 59 per cent, while the generalization to the other word categories was *the greater the higher the frequency of such word categories in free association tests and the faster their reaction time in controlled association tests* (61, 32). However, this relationship held only if the relatedness between the conditioned and the generalization words was classified according to "relatedness of conditioned words to generalization words," and not as a "relatedness of generalization words to conditioned words"; e.g., only if a conditioned word of "dog" and a generalization word of "animal" or a conditioned word of "flower" and a generalization word of "petal" were classed respectively as subordinates and as part-whole's, and not as supraordinates and as whole-part's. The significance of this finding will be discussed in a later section.

Effects of Mental Sets upon Semantic CR Generalization (50)

This study was largely an extension of the one preceding. Before being tested for generalization to various word categories, nine subjects were divided into three equal and equated groups. Group C_1 practiced *corresponding* word categories in controlled association tests, e.g., practiced supraordinates before being tested for CR generalization to supraordinates. Group C_2 practiced *converse* word categories, e.g., practiced subordinates before CR tests for supraordinates, while Group C_3 was not subjected to any such preliminary practice. As a result, Group C_3, the control group, showed a mean CR generalization of 34 per cent. Group C_1 that practiced corresponding word categories manifested a mean generalization of 54 per cent, and Group C_2 that practiced converse categories, a mean of 25 per cent. The t's here are significant for all three group differences with P equalling .01.

Effects of Cognitive (Knowledge of Stimulus Relations), Voluntary-Facilitatory, and Voluntary-Inhibitory Attitudes

Twelve subjects, divided into four equal and equated groups, were, first, conditioned to the tone C on a Stoelting *just* harmonium and to the word *flower;* and, then, tested for generalization to (a) tones F, B, e, a, d^1, g^1, $c\sharp^2$, $f\sharp^2$, and to (b) words "flour," "glower," "shower," and "scour." The cognitive attitudes for tone generalization were induced by the instructions of: "When you hear a tone in this session, try to think of its relatedness to the tone with which we experimented in the last two sessions. Try to think whether the tone that you will hear is higher or lower and how much higher and how much lower than the tone with which we experimented." The voluntary-facilitatory attitudes were instilled by: "I'd like to see how well you can control reactions. You were conditioned to secrete saliva to a low tone on this instrument [conditioning explained at some length]. From now on, you will hear eight higher tones, each about two or three musical intervals higher than the other, and it is your task to secrete most saliva to the tone closest to the conditioned tone—less, though, than to the conditioned tone itself—and least saliva to the tone most removed from the conditioned tone [all eight tones are now demonstrated]. In other words, I would like you to build up a saliva scale that will correspond as closely as possible to the scale of tones here." The instructions for the voluntary-inhibitory attitudes differed from those for the voluntary-facilitatory attitudes in that the subjects were told to "secrete an equal amount of saliva to each tone irrespective of its pitch or quality." The instructions for word generalizations were similar to those of tone generalization except that the subjects were also told "not to pay attention to word relatedness in meaning but only to that of sound and spelling." Finally, there was a control group whose subjects were not given any instructions about the generalization.

The results are presented in Table VI. Space forbids a detailed discussion of the table. But a summary statement can be made: *within the limits of the present experiment cognitive and voluntary attitudes modified only a little the subjects' gradients of CR generalization.*

TABLE VI

EFFECTS OF COGNITIVE (KNOWLEDGE OF STIMULUS RELATIONS), VOLUNTARY-FACILITATORY, AND VOLUNTARY-INHIBITORY ATTITUDES UPON STIMULUS GENERALIZATION OF SALIVARY CONDITIONING IN HUMAN SUBJECTS

The subjects were 12 college undergraduates and the conditioned stimuli were (a) the tone *C* on a harmonium and (b) the word *flower*. Each entry is a mean of six determinations, two from each of the three subjects. VF = voluntary-facilitatory attitude, VI = voluntary-inhibitory, KR = knowledge of stimulus relations, UI = uninstructed.

Induced Attitude	*Per Cent of Generalization* *Generalization Stimuli* *Tones*								*Reliable Gradient Steps (estimated)*
	F	*B*	*e*	*a*	d^1	g^1	$c\sharp^2$	$f\sharp^2$	
UI	67	49	54	42	40	32	41	38	Three
VF	64	52	44	50	39	40	30	24	Three or Four
VI	78	72	76	74	56	58	42	48	Two or Three
KR	62	54	50	41	34	40	32	31	Three or Four

Words

	Flour	*Glower*	*Shower*	*Scour*	
UI	39	29	32	19	Three
VF	62	68	52	23	Three
VI	68	72	60	48	Three
KR	54	41	45	26	Three

The Experiment with Transliterated Russian Words

Nine subjects, unfamiliar with Russian, were divided into three equal and equated groups and conditioned to four transliterated Russian words: BUKVA, DOLGO, KUPIT, and SMESHNOY. Group N was not told the meanings of the words. Group G was told the meanings of the words before being tested for generalization, while Group CG was given one set of meanings before the primary training of the CR, and another reverse set of meanings before the generalization tests (having been assured that the second set was the correct one). The subjects in Group N did not, of course, show any significant semantic CR generalization. But the important point brought out by the experiment was that the course and range of the semantic generalization was determined entirely by the meanings given to the Russian words before the generalization tests. *The meanings imparted prior to the original training of the CR had no effect upon the subsequent semantic generalization of the CR to the verbal stimuli.*

Summary of Razran's Results

Besides corroborating in the main the findings of the Yale and the Pavlov laboratories (as analyzed by the writer), the writer's results indicate that:

1. With human subjects, gradients of CR generalization, though very crude ones, are more evident with verbal than with sensory conditioned stimuli.
2. With musical intervals as the conditioned stimuli, common ratios determine much more the course of CR generalization than common constituent tones, either fundamentals or partials; with flashes of lights as the conditioned stimuli, the spatial patterns of the lights are more determining than the colors of the lights in the patterns; and with verbal conditioned stimuli, relatedness in meanings is more significant than relatedness in sound and spelling.
3. Mental sets are effective in determining the general direction and magnitude of CR generalization but do not seem to be of much significance—without special training—in creating or in counteracting the *gradients* of the generalization.
4. The course and very existence of CR generalization is more likely a function of the subsequent testing for generalization than of the original training of the conditioning.[8]

The Total Evidence and the Lashley-Wade Interpretations

Total evidence adduced heretofore is, in the writer's estimate, discordant with either (a) the views of Pavlov that CR generalization is a function of the undulation of cortical excitation or (b) the views of Hull that the generalization gradient is a monotonic decreasing function of the magnitude of the differences between the conditioned and the generalization stimuli. But the possible concordance of the evidence with the Lashley and Wade interpretations—outlined in the Introduction—deserves careful examination.

---

[8] In one study by the writer (50), the CR generalizations of four subjects, who were conditioned to controversial sentences such as "socialism is desirable," seemed to be affected by the socio-political views of the subjects, while the conditioning itself showed no such differences. However, these results were obtained from a very small number of subjects and must be considered very tentative.

Generalization as a failure of association. There is no doubt in the writer's mind that a good deal of presumed evidence of CR generalization is due to a failure of association. To Lashley and Wade's insightful observations, he would add the following experimental support:

1. In six different experiments from Pavlov's laboratory (24, 27, 37, 53, 62, 63) it was shown that dogs conditioned to a compound stimulus—light and taction, light and sound, two different sounds or lights or tactile stimuli—had usually failed to produce a CR, when only one of the stimuli, supposedly the weaker one, was presented alone. There was no generalization to these components of the compound stimulus, even though they were physically present during the original CR training and even though they obviously were physically similar to the compound stimulus (partial identity.) And the only sensible interpretation of this phenomenon would seem to be that the animals failed to associate these particular component stimuli with the total CR situation, that the stimuli were not "attended to," were present but ineffective. Pavlovian explanations that in such cases the weaker stimuli become inhibited by the stronger ones contradict their own doctrine that all aspects of a stimulus are associated with the conditioned response. Or, to put it differently, if weaker stimuli are inhibited by stronger stimuli during primary conditioning, why do not the weaker generalization bonds become inhibited by the stronger conditioned bonds, and why is there generalization altogether?
2. In a number of experiments from Pavlov's laboratory, partially decorticated or badly injured or specially drugged dogs are often reported to manifest a great deal of CR generalization. And a widening of generalization is also often described in old dogs, young puppies, and in lower animals. Inasmuch as in all these case associative capacity is no doubt reduced, an interpretation of their generalizations in terms of failure of association seems very reasonable.

However, against these two considerations the writer would pit four others: viz.:

1. Partially decorticated and injured, and old dogs have also been reported by Pavlov and his students as having reduced greatly the scope of their generalization.
2. The frequent increase of CR generalization in the initial stages of training the CR—which the writer believes to be an established fact—does not go well with a "failure of association" interpretation.
3. Failure of association could hardly apply to the generalizations obtained by Grandine and Harlow (9) and by Grice (10) who used a Lashley-Wade technique.
4. Failure of association could not explain the generalizations in the writer's experiments in which the subjects attended very minutely to the stimuli, and probably not too well the generalizations in other experiments with human subjects.

In other words, while failure of association may account, in the writer's estimate, for a good deal of *reported* CR generalization, it does not explain *all* CR generalization reported. Probably Lashley and Wade did not themselves mean it to do so.

Generalization as a product of variable stimulus thresholds. Again, it seems well probable that some data on CR generalization are to be attributed to variable stimulus thresholds. In spatial generalization of taction, Loucks (30) actually demonstrated what he called peripheral mechanical irradiation by stimulating a point not too far from the conditioned spot with a "pricker" obtained from Pavlov's laboratory. And the writer pointed out earlier in this article the probable masking effects of intensity in CR generalization. Still, there is a limit to variable thresholds, and the writer doubts very much whether such a concept could be of any value in interpreting CR generalization to stimuli far removed from the conditioned stimulus, inter-dimensional and inter-sensory generalization, generalization by the Lashley-Wade technique, and generalization in the writer's experiments.

Generalization along discriminated dimensions. The main argument here is the Lashley and Wade statement, quoted earlier, that "It [true generalization] does not occur in conditioning to a single stimulus but is somehow a function of differential training with two or more stimuli on the same dimension" (p. 82). This apparently means that, according to Lashley and Wade, whatever CR generalization cannot be explained by "failure of association" or by "variable stimulus thresholds," its cause must be sought in the reactional biographies of the stimuli involved, a laudable but difficult matter. To be sure, reactional biographies are very important, and the writer could point to the experiments by Prokofiev and Zeliony (38) and particularly to the one by Brogden (6) on sensory pre-conditioning which show the interactions of sensory stimuli in animals without even reinforcement; as well as to the effects of previous experience on CR gradients in his own experiments. Yet in the 67 studies from Pavlov's laboratory which the writer analyzed he ruled out results with dogs that had had, as far as could be ascertained, some previous experience with the generalization stimuli. And in the writer's own experiments, discriminative sets modified but did not wholly change the character of the subjects' CR gradients. Furthermore, on general grounds and analogies, Lashley and Wade's total denial of the existence of a true single-stimulus absolute value generalization is not wholly borne out by their own statement. They state: "... memory for relations is much more permanent than memory for absolute properties" (p. 86), which to the writer means that transposition is more significant than generalization, a view with which the writer tends, in a large way, to agree (cf. 44). But to the writer the statement also means an admission of the existence of memory for absolute properties and, by analogy, the existence of a true single-stimulus CR generalization.[9]

The Writer's Own Interpretations

The writer's own interpretations of CR generalization rest upon three considerations: *a bifactoral theory; a "subsequent testing" hypothesis; a doctrine of a qualitative, categorizing, "rating scale" type CR gradient.*

The bifactoral theory. The bifactoral theory of CR generalization maintains that there are two kinds of generalizations: (a) *pseudo-generalization* and (b) *true generalization*. Pseudo-generalization, well accounted for by Lashley and Wade's interpretations, correlates, for the most part, negatively with organismic capacity. It predominates in the CR generalizations of lower animals (fish, guinea-pigs), in partially decorticated or very young or very old higher animals, in states of fatigue and inattention, and in some stages of CR training (initial, over-training). True generalization, on the other hand, is a positive capacity of the organism, an ability to generalize *absolute* characteristics of stimuli or objects. It seems to increase with maturity, alertness, post-operative recovery from cortical extirpations, administration of some drugs, and other capacity-enhancing influences.[10] Analogically, perhaps, the relation of pseudo-generalization to true generalization is not unlike that of the undifferentiated total action of the young foetus to the structured whole activites of the fully developed individual.

The subsequent testing hypothesis. The "subsequent testing" hypothesis asserts that CR generalization develops, not during the original training of the conditioned stimuli, but during the subsequent testing of the generalization stimuli. The writer agrees here fully with Lashley and Wade that "there is no 'irradiation' or spread of effects during primary conditioning" (p. 74) and disagrees with the assumptions of Pavlov and Hull and Spence that during original CR training some sort of generalization bonds, latent but later ready to function, are automatically developed. He believes that such assumptions are not only cumbersome physiologically and superfluous logically, but also that they are contradicted by a series of his own experiments. According to his view, *all effects of generalization are generated during tests of generalization*.[11]

---

[9] The writer agrees also with Lashley and Wade that the "basic nervous mechanism is one of reaction to ratios of excitation" (p. 86). But he believes that some ratio of excitation is also involved in simple conditioning, a ratio between the reaction-patterns of the conditioned and the unconditioned stimuli.

[10] The writer regrets not to be able to discuss here the evidence for these assertions. One reason is lack of space. Another is that the evidence has not been fully analyzed yet, and that these assertions are thus based only upon impressions of reading Russian studies, and in the case of lower animals upon an earlier, not completely up to date, review by the writer (**41**) of the topic.

The doctrine of a qualitative, categorizing, "rating scale" type of CR gradient. This doctrine is in a way the crux of the writer's interpretations. It is based upon a statistical and a logical analysis of total evidence which, in the writer's estimate, demonstrates that there is a true CR gradient, but that this gradient is very qualitative and very crude, consisting of only a few steps, perhaps more steps in human beings than in dogs, but few just the same. Apparently, when human beings or dogs that have been conditioned to some stimulus or object are confronted with some new non-conditioned but in some way related stimulus or object, they categorize or rate the new stimulus on some sort of crude similarity-dissimilarity scale. With human subjects, introspections actually reveal such categorizing attitudes as "similar," "very similar," "not so similar," "somewhat similar," "dissimilar," "very dissimilar," and the like—attitudes that apparently control or even initiate the generalization responses. And it is the writer's contention that some such categorizing behavior is operative also in animals.

This categorizing is thought by the writer to be very dynamic and changeable and *varying much more with the organic dimensions of the organism than with the physical dimensions of the external stimuli in the CR situation* (data deny significant variations with the latter). Its neurophysiology is regretfully obscure, as regretfully as the neurophysiology of learning in general.[12] But it should not be difficult to conceive the organic dimensions with which it might correlate: perhaps the movement-produced stimuli of Guthrie, or the situation-sets of Woodworth, or even the means-ends-capacities of Tolman. At any rate, qualitative step-like CR gradients and only such gradients are objectively demonstrable. So that those of the writer's colleagues who fear anthropomorphism more than zoomorphism need accept only the type of gradient and not its human analogy. While to others, *rating scales in dogs* should be no less welcome than *hypotheses in rats*.

GENERAL SUMMARY

Four views of CR generalization have been considered:

1. *Cortico-physiological* (Pavlov, Bekhterev): CR generalization is a function of a wave-like irradiation of cortical excitation. It posits a quantitative continuous CR gradient.

2. *Physico-behavioral* (Hull, Spence): CR generalization is a monotonic de-

---

[11] It might be worth considering a similar hypothesis for transposition and for more complex transfer phenomena.

[12] The writer is much more in accord with Loucks than with Skinner and Hull about the need and value of correlating behavior with neurophysiology; but he does defer to Lashley about the paucity of our knowledge in this area.

creasing function of the magnitude of the differences between the conditioned stimulus and the generalization stimuli. It, too, posits a quantitative continuous CR gradient.

3. *Failure of association-transposition* (Lashley and Wade): some reported data on CR generalization are due to failure of association; others represent transpositions of discriminated dimensions. It denies the existence of true CR generalization.

4. *Categorizing-rating* (Razran): While recognizing the roles of failure of association and transposition, it affirms the existence of a true CR generalization which is attributed to the organism's categorizing or rating of related stimuli on some sort of crude similarity-dissimilarity scale. It posits a very crude and qualitative CR gradient.

Adduced evidence favors the fourth view and is clearly in disagreement with the first two views.[13]

BIBLIOGRAPHY

1. ANREP, G. V. The irradiation of conditioned reflexes. *Proc. Roy. Soc. London*, 1923, **94B**, 404–425.
2. BASS, M. J., & HULL, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, **17**, 47–65.
3. BEKHTEREV, V. M. *Psychology, reflexology, and Marxism.* Leningrad: GIZ, 1925.
4. BEKHTEREV, V. M. *General principles of human reflexology.* London: Jarrolds, 1933.
5. BERITOFF, J. S. *Individually-acquired activity of the central nervous system.* Tbilisi: GIZ, 1932.
6. BROGDEN, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.
7. BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.
8. CARR, H. A. The laws of association. *Psychol. Rev.*, 1931, **38**, 212–228.
9. GRANDINE, L., & HARLOW, H. F. Generalization of the characteristics of a single learned stimulus by monkeys. *J. comp. Psychol.*, 1948, **41**, 327–338.
10. GRICE, C. R. The effect of differential training to single stimuli upon the acquisition of a size discrimination habit. *Amer. Psychologist*, 1948, **3**, 242.
11. GUTHRIE, E. R. Conditioning as a principle of learning. *Psychol. Rev.*, 1930, **37**, 412–428.
12. GUTHRIE, E. R. Pavlov's theory of conditioning. *Psychol. Rev.*, 1934, **41**, 199–206.
13. GUTHRIE, E. R. *The psychology of learning.* New York: Harpers, 1935.
14. HILGARD, E. R. *Theories of learning.* New York: Appleton-Century, 1948.
15. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
16. HOVLAND, C. I. The generalization of

---

[13] The writer regrets the omission, through an oversight, of the pertinent study by Grant and Dittmer (*J. exp. Psychol.*, 1940, **26**, 299–310). However, only 2 of the 31 subjects in the study showed generalization gradients without reversals while 17 subjects had mean CRs to generalization stimuli greater than to the conditioned stimulus, so that the results of these experimenters are in line with the present analysis. Grant himself is of course by no means an adherent of Pavlovian irradiation and—the writer ventures—of Hullian generalization.

conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.

17. Hovland, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.
18. Hovland, C. I. The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1937, **21**, 261–276.
19. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
20. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
21. Hull, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120–134.
22. Hull, C. L., *et al. Mathematico-deductive theory of rote learning.* New Haven: Yale Univ. Press, 1940.
23. Humphreys, L. G. Generalization as a function of the method of reinforcement. *J. exp. Psychol.*, 1939, **25**, 141–158.
24. Kashernininova, N. A. *Contributions to the study of salivary conditioned reflexes in response to tactile stimuli in the dog.* Thesis, St. Petersburg, 1908.
25. Konorski, J. *Conditioned reflexes and neuron organization.* Cambridge: Cambridge Univ. Press, 1948.
26. Krasnogorski, N. I. Bedingte und unbedingte Reflexe in Kindesalter und ihre Bedeutung für die Klinik. *Ergebn. inner. Mediz. Kinderhk.*, 1931, **39**, 613–730.
27. Kupalov, P. S., & Gantt, W. H. The relation between intensities of conditioned stimuli and magnitudes of conditioned reflexes. *Trud. Fiziol Lab. Pavlova*, 1928, **2**, 1–12.
28. Lashley, K. S. Learning: I. Nervous mechanisms in learning. In C. Murchison (Ed.), *The foundations of experimental psychology.* Worcester, Mass.: Clark Univ. Press, 1929. Pp. 524–563.
29. Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
30. Loucks, R. B. An appraisal of Pavlov's systematization of behavior from the experimental standpoint. *J. comp. Psychol.*, 1933, **15**, 1–45.
31. Loucks, R. B. Reflexology and the psychobiological approach. *Psychol. Rev.*, 1937, **44**, 320–338.
32. May, M. A. The mechanism of controlled association. *Arch Psychol., N. Y.*, 1917, No. 39.
33. Pavlov, I. P. *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex.* London: Oxford University Press, 1927.
34. Pavlov, I. P. *Lectures on conditioned reflexes: twenty-five years of objective study of the higher nervous activity (behavior) of animals.* New York: International Publishers, 1928.
35. Pavlov, I. P. The reply of a physiologist to psychologists. *Psychol. Rev.*, 1932, **39**, 91–127.
36. Pavlov, I. P. *Conditioned reflexes and psychiatry.* New York: International Publishers, 1941.
37. Perelzweig, I. *Materials to the study of conditioned reflexes.* Thesis, St. Petersburg, 1907.
38. Prokofiev, G., & Zeliony, G. Des modes d'associations cérébrales chez l'homme et chez les animaux. *J. de Psychol.*, 1926, **23**, 1020–1028.
39. Razran, G. Theory of conditioning and related phenomena. *Psychol. Rev.*, 1930, **37**, 25–43.
40. Razran, G. Conditioned responses in

children: a behavioral and quantitative critical review of experimental studies. *Arch. Psychol. N.Y.*, 1933, **23**, No. 148.

41. Razran, G. Conditioned responses in animals other than dogs: a behavioral and quantitative critical review of experimental studies. *Psychol. Bull*, 1933, **30**, 261–324.

42. Razran, G. Conditioned responses: an experimental study and a theoretical analysis. *Arch. Psychol., N.Y.*, 1935, **28**, No. 191.

43. Razran, G. Transposition of relational responses and generalization of conditioned responses. *Psychol. Rev.*, 1938, **45**, 532–538.

44. Razran, G. Studies in configural conditioning: VII. Ratios and elements in salivary conditioning to various musical intervals. *Psychol. Rec.*, 1938, **2**, 370–376.

45. Razran, G. The nature of the extinctive process. *Psychol. Rev.*, 1939, **46**, 264–297.

46. Razran, G. A simple technique for controlling subjective attitudes in salivary conditioning of adult human subjects. *Science*, 1939, **89**, 160–161.

47. Razran, G. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.

48. Razran, G. Studies in configural conditioning: V. Generalization and transposition. *J. genet. Psychol.*, 1940, **56**, 3–11.

49. Razran, G. Semantic and phonetographic generalizations of salivary conditioning to verbal stimuli. *J. exp. Psychol.* (Oct. 1949, in press).

50. Razran, G. Sentential and propositional generalizations of salivary conditioning to verbal stimuli. *Science*, 1949, **109**, 447–448.

51. Razran, G. Some psychological factors in the generalization of salivary conditioning to verbal stimuli. *Amer. J. Psychol.*, 1949, **62**, 247–256.

52. Razran, G., & Warden, C. J. The sensory capacity of the dog as studied by the conditioned reflex method (Russian schools). *Psychol. Bull.*, 1929, **26**, 202–222.

53. Rickman, V. The problem of the intensity of conditioned reflexes. *Trud. Fiziol. Lab. Pavlova*, 1928, **2**, 13–21.

54. Robinson, E. S. *Association theory today.* New York: Appleton-Century, 1932.

55. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

56. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

57. Spence, K. W. Continuous versus non-continuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **47**, 271–288.

58. Warren, H. C. *A history of the association psychology.* New York; Scribners, 1921.

59. Watson, J. B., & Raynor, R. Conditioned emotional reactions. *J. exp. Psychol.*, 1920, **3**, 1–14.

60. Wickens, D. D. Studies of response generalization in conditioning: I. Stimulus generalization during response generalization. *J. exp. Psychol.*, 1943, **33**, 221–227.

61. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

62. Zeliony, G. *Reactions of dogs to auditory stimuli.* Thesis, St. Petersburg, 1907.

63. Zeliony, G. Contribution a l'analyze des excitants complexes des reflexes conditionnels. *Arkh. biol. Nauk*, 1910, **15**, 437–453.

64. Trudy Fiziologicheskikh Laboratorii Akademika I.P. Pavlova, 1924–1941, 1–10.

Received March 6, 1949.

PSYCHOTHERAPY AS A PROBLEM IN LEARNING THEORY[1]

EDWARD JOSEPH SHOBEN, JR.
State University of Iowa

It has become increasingly apparent that clinical psychologists are more and more drawing psychotherapy into their compass of activities. If this enlargement of scope is to be something more than a trading of one's psychological birthright for a share of psychiatric pottage, it would seem imperative that the therapeutic functions of the psychologist be regarded from the point of view of research as well as from that of practice. As Sanford (47) puts it,

> What should be of great help to us here is our training in scientific method and our tradition of research-mindedness. It would be hard to name an area in which research is more needed than it is in therapy, or an area in which what is being done lags further behind what might be done. . . . And one might say, furthermore, that it is primarily up to the psychologist to perform this needed research.

The difficulties in the way of such inquiry, however, are enormous, as is well attested to by the paucity of investigations of the therapeutic process in terms of the problems, techniques, and concepts common to general psychology. The nature of some of these barriers to psychological research on a matter of such importance probably merits some brief attention.

In the first place, there are situational deterrents to research in psychotherapy. Counseling[2] usually takes place in a "service" setting and is consequently seldom subject to the kinds of exact manipulation required by rigorous experimentation. Often, attempts to control various factors in the therapeutic set-up give rise to serious ethical problems concerning the relationship of the therapist and his agency to their clients, and certainly the pressure of the demand for counseling services frequently conflicts with the requirements of a research program. Secondly, the problem of complexity gives one pause. Psychotherapy is a form of social interaction, an active social situation, in which many subtle, difficult-to-isolate aspects of the personalities of *both* patient and counselor must be taken into consideration. The therapist is not merely the wielder of some supposedly meliorative technique but is deeply involved as a personality in the counseling process. Thus, the psychology of the psychologist, as well as the psychology of the patient and the nature of the therapeutic method, enters into the determination of the therapeutic end product. Third, there are personnel problems militating against effective research in psychotherapy. Psychologists most familiar with the therapeutic process are seldom well schooled in the experimental and conceptual skills basic to fruitful investigations in general psychology, whereas those who are best equipped technically and conceptually as research workers are generally rather untutored in therapeutic techniques, are unfamiliar with clinical material, and are frequently repelled by the admittedly gross and somewhat nebulous notions clinicians use in their efforts to conceptualize the complex phenomena with which they work. In sum, the situational lack of amenability of psychotherapy to experimental inquiry, the enormous complexity of the factors entering into the counseling process, and the differences in training and interest between clinical and laboratory workers all tend to impede a *rapprochement* between psychotherapy and the research functions characteristic of general psychology.

---

[1] This article represents a revision and extension of an earlier attempt (51) to conceptualize psychotherapy in terms of systematic behavior theory. Acknowledgment must be made to a number of people, foremost among whom is Dr. O. H. Mowrer, who, though he may recognize some of his ideas in the ensuing pages, must not be held responsible either for their form or for the uses to which they are put. Others are Dr. Kenneth Spence and Dr. I. E. Farber, of the University of Iowa, who have been invaluable sources of stimulation and instruction but who are absolved from any responsibility for what is here said.

[2] The terms counseling and psychotherapy are here used interchangeably without regard for any of the distinctions they are sometimes employed to convey.

In spite of these difficulties, there is one slender lead that might be profitably followed in the attempt to provide a basis for the conceptualization and investigation of psychotherapy as a problem in general psychology. This is the widespread recognition that psychotherapy is essentially a learning process and should be subject to study as such.

This point of view is not only in harmony with the general conception of counseling as a conversation or series of conversations between two persons, therapist and patient, the goal of which is to resolve the conflicts, reduce the anxiety, or somehow modify the behavior of the latter—a conception which clearly implies learning; it has been more or less clearly so verbalized by a number of clinical workers. Cameron (4) sees the desideratum of counseling as the patient's "acquisition of normal biosocial behavior," a statement which definitely implies the learning of new ways of reacting as a function of the therapeutic process. Alexander and French (1) advance as a basic therapeutic principle the reexposure of the client, within the favorable circumstances of psychotherapy, to emotional situations with which he was unable to deal in the past. Presumably, the justification for such a reexposure rests on the hypothesis that its occurrence "under more favorable conditions" in some way permits the patient to learn more adequate ways of coping with such experiences. Rogers (45) describes the therapeutic process as a freeing of the "growth capacities" of the individual which permits him to acquire "more mature" ways of reacting. If "growth" in this context means (as it must) something more than physiological maturation, and if it is not to be lumped with the old and rather mystic homeopathic notion of the *vis medicatrix naturae*, it must refer to the client's acquisition of new modes of response. Such new modes of response are "more mature" because for a given patient they are less fraught with anxiety or conflict. Thus, Rogers is actually talking about psychotherapy as a learning process. White (64) insists that, "Psychotherapy is designed to bring about learning . . . "; and Darley (7) argues that unless the process of learning in counseling is demonstrated, it is not legitimate to infer that the modifications of behavior that may occur during or following therapy are necessarily outcomes of therapy.

In spite of this widespread acknowledgement of psychotherapy as a learning process, there have been few attempts (11, 25, 49, 50) systematically to formulate therapy in terms of learning theory. This paper represents a tentative, apologetically offered effort to construct a learning-theory interpretation of counseling that will help to narrow the gap between practitioner and researcher, clinician and experimentalist, and to encourage some much needed investigation.

COMMON FACTORS IN SCHOOLS OF PSYCHOTHERAPY

When one surveys the various theories and practices of psychotherapy in an effort to find those common factors which a learning-theory interpretation of the counseling process must cover, it appears possible to make four summarizing general statements:

1. All schools of psychotherapy can with some justice claim cures (46). Notable successes seem to be the common property of virtually all forms of counseling from moral suasion through non-directive therapy to psychoanalysis.
2. Clinical patients,[3] in spite of their enormous differences, tend to present a similar problem in that one of their primary motivations is anxiety and much of their non-integrative or "symptomatic" behavior is maintained on the basis of anxiety reduction.
3. The goal common to most psychotherapies is the modification of the client's underlying anxiety. This is related to the hypothesis that once his motivation is altered, the overt habit structures of the patient will change.

---

[3] The "clinical patients" spoken of in this paper include only those classifiable as neurotic or "maladjusted." Nothing said here is meant to apply to psychotics, psychopaths, or behavior problems associated with endocrine disturbances or lesions of the central or autonomic nervous systems.

4. Finally, all types of counseling employ the techniques of the *therapeutic relationship*, the unique social situation that is formed when therapist and patient meet to discuss the problems of the latter, and of *conversational content*, that is, of talking about certain things within the therapeutic setting rather than others.

A word must be said about each of these four factors which seem to be common to the various forms of counseling, regardless of the doctrinal banners flown.

All schools report cures. If it is true that the proponents of various theories of psychotherapy all seem able to claim successes, and if it is true—as has often been pointed out—that successes are no proof of therapeutic theory, then it would seem to follow that an understanding of the counseling process would be furthered by giving more attention to the conditions under which the patient's learning of new modes of reaction takes place within the general clinical setting. If this is a fair notion, based as it is on the conception of therapy as a learning situation, it might be instructive to explore the points in common among the different approaches to counseling in terms of (a) the similarity of patients' problems, (b) the agreement among clinicians as to goals, and (c) the techniques common to nearly all therapeutic enterprises. Such an exploration might lead to a formulation of the learning process in counseling in terms of these three sets of information.

Similarities in clinical cases. While from the practical standpoint of dealing therapeutically with patients it is necessary to consider each case in all its uniqueness, from a theoretical point of view it is instructive to look for similarities. This amounts to asking the rather ambitious questions of (a) What constitutes the core of "neurosis" or "maladjustment"? and (b) What are the common problems faced by therapists in their contacts with patients? While no definitive answer can be given here, it is important to consider these issues as bearing on the goals and techniques employed by counselors of different theoretical persuasions and as factors to be accounted for in attempting to formulate a learning-theory interpretation of the therapeutic process.

A point on which there seems to be widespread agreement is, in Horney's (17) phrase, that "one essential factor common to all neuroses . . . is anxieties and the defenses built up against them." The phenomena clinically identified as feelings of insecurity, feelings of inadequacy, and guilt feelings are all variants of anxiety in the sense that they involve debilitating expectations of future punishment. Likewise, it would seem that the "phenomenological self-concept" of Combs (6) and Rogers (45) refers to little more than a patient's level of anxiety, guilt, or inadequacy, together with his verbalizations, accurate or otherwise, of his defenses against them.

To conceptualize anxiety usefully, it is necessary to discriminate between anxiety and fear or, as Freud (12, 13, 14) did, between neurotic anxiety and objective anxiety. Fear may be thought of as an affective reaction proportionate to some external danger. Anxiety, on the other hand, differs from fear in at least two ways. First, if one asks a "neurotic" patient what he is afraid of, he will admit to being afraid but will generally have no idea of what the source of the possible danger might be. Anxiety may be aptly termed either a fear of "nothing" or a fear of something which is objectively irrelevant. Second, while both fear and anxiety are anticipatory states involving some kind of premonition of danger, the signal to which anxiety is a reaction is usually internal, some impulse to act in a way that has been forbidden. An illustrative case may clarify this point.

E. B., a 24-year-old male undergraduate veteran, despite slightly better than average academic ability, is making poor grades and is in danger of being dismissed from his university. He complains of being "unable" to study, feelings of inferiority in social groups, and serious doubts as to both his intellectual and social adequacy. He has some guilt feelings about having transferred from a premedical curriculum to English, because his parents are quite eager for him to become a physician. His father is a farmer who has been quite successful financially and in community politics, and who has been highly ambitious for his son. He has imposed very high standards of attainment on the boy, has been quite strict and stern with him and has had a number of set ideas which he felt that the youngster should accept and act upon "for his own good." Any deviation on the part of the patient from the parentally prescribed ways of doing was met with severe punishment, the verbal part of which usually consisted in a variety of changes rung on the theme of the boy's worthlessness and a series of predictions that he would come to no good end. In short, any self-initiated activity—behavior which the parents themselves did not lay out—was fraught with danger. When the boy began counseling, he was squarely on the horns of a dilemma: unable to meet parental demands for a variety of reasons, he was also unable to initiate any divergent plans of his own without experiencing a flood of anxiety, i.e., anticipations of parental punishment.

This, if it is acceptable, leads to a general formulation of nonintegrative or neurotic behavior. Anxiety has repeatedly been shown to have drive properties (**29, 32**), and on the basis of the anxiety drive, individuals who are maladjusted seem to develop various overt reaction patterns that become stable according to the degree to which they reduce the anxiety. This statement in terms of contemporary reinforcement theory (**19**) is quite in keeping with Freud's (**12**) idea of the interchangeability of anxiety and symptom, by which he means that through the formation of symptoms the patient protects himself from anxiety attacks. Anxiety is allayed by some anxiety-reducing symptom; if the symptomatic behavior is somehow eliminated, the anxiety returns. On the basis of this notion it is possible to define a neurosis or a maladjustment in terms of behavior which serves to reduce anxiety directly *without altering the conditions which produce the anxiety.* Freud consistently refers to anxiety as a signal of impending danger; the maladjusted person is one who either consciously or unconsciously engages in acts which eliminate or neutralize the signal while leaving the objective danger unaffected. He is in the position of the motorist who shuts his eyes to warnings of dangerous curves, thus protecting himself from worry but leaving himself liable to serious accidents.

Such a conception permits an explanation of the curious observation that non-integrative behavior is at the same time self-defeating and self-perpetuating. It is self-defeating in that such behavior leads inevitably to further punishment: the motorist has accidents; the illustrative case suffers academic failures and social disarticulation through his avoidance of study to protect himself from the anxiety engendered by self-initiated activity and his withdrawal from social affairs to hide his "worthlessness." It is self-perpetuating because of the immediate reinforcement derived from anxiety reduction. Since the occurrence of a reinforcing state of affairs lies on the temporal gradient of reinforcement in greater proximity to the anxiety-reducing behavior than does the more remote punishment, the connection between the external and internal cues of anxiety and the non-integrative response tends to be strengthened (38).

A necessary concept in a theory of anxiety is that of repression. This notion refers to the exclusion from communicability (consciousness) of an impulse to act which has led to punishment. When a parent punishes a child severely for some tabooed act, the impulse to commit such an act becomes, through its association with the punishment, a stimulus for anxiety. One way by which the anxiety may be avoided is through repression—the exclusion from awareness of the impulse. If the repression is complete, there is a thorough-going allaying of anxiety, and the forbidden impulse no longer constitutes a problem.

Difficulty arises because repression is seldom if ever complete. The individual is constantly threatened by "a return of the repressed" (14) which touches off anxiety without the patient's being able to verbalize the cues for it. In short, the repressed impulse, although excluded from communicability, is still operative at subliminal levels. Why this should be true is something of a psychological mystery, although some light is shed upon it by investigations of punishment. Estes (9), for example, by a series of experiments has shown that punishment does not extinguish a response which has been positively reinforced. He concludes,

> . . . a response cannot be eliminated from an organism's repertoire more rapidly with the aid of punishment than without it. In fact, severe punishment may have precisely the opposite effect. . . . The punished response continues to exist in the organism's repertoire with most of its original latent strength. While it is suppressed, the response is not only protected from extinction, but it also may become a source of conflict. An emotional state, such as "anxiety" or "dread," which has become conditioned to the incipient movements of making the re-

> sponse will be aroused by any stimuli which formerly acted as occasions for the occurrence of the response (pp. 37–38).

This provides a neat parallel to what is implied in the concept of repression.

In summary, then, one might say that clinical cases share in common (a) anxiety touched off by (b) unverbalized, unsuccessfully repressed impulses to act in ways that have met with punishment, and (c) persistent non-integrative behavior of many kinds, which reduces the anxiety but does nothing about eliminating its objective causes.

Common goals in psychotherapy. In spite of its non-integrative nature, overt neurotic behavior acquires remarkable persistence through anxiety-avoidance. This persistence is probably the factor most responsible for the failure and consequent elimination of clinical techniques aimed at the elimination of symptoms. Such a goal, in effect, defined psychotherapy as a process of robbing the patient of his defenses against anxiety without alleviating the unbearable state of dread. Since such an end is impossible of realization, advice, persuasion, exhortation, and suggestion have largely gone by the board in favor of methods which focus on the client's anxiety itself.

In other words, the goal of most modern psychotherapies is the modification of the emotional determinants of neurotic behavior. Thus, Alexander and French (1) speak of therapy as "a corrective emotional experience," which presumably results in a diminution of anxiety and a consequent elimination of persistent non-integrative behavior from the patient's repertoire. Likewise, White (64) points out that "Psychotherapy does not take place primarily in the sphere of intellect . . . . Its sphere of operation is the patient's feelings." The kind of learning with which counseling is concerned has to do chiefly with the alteration of motives and affective drives. This does not mean, of course, that the therapist is uninterested in his client's overt behavior; on the contrary, it is his job to help the patient alter it and achieve a repertoire of more integrative habits. But since this goal does not seem attainable through any kind of direct manipulation, the counselor generally works on the elimination of the basic anxieties, implicitly hypothesizing that once the drive conditions are changed, the neurotic behavior will show less strength.

Common tools in psychotherapy. From the standpoint of technique, there are two main aspects of the counseling process, common to all schools of psychotherapy. One is the unique *relationship* that develops between therapist and patient; the other is the *conversational content*, what they talk about during their sessions together. The proponents of different theories of counseling may emphasize one or the other of these factors, but both figure in their final formulations of therapeutic procedure. Thus, Williamson (**66**) and Kraines (**24**) stress the therapist's obtaining personal information from the client so that the counselor may guide him somehow to a higher level of adjustment. In spite of this emphasis, both these clinicians devote a good deal of attention to the necessity of establishing and maintaining rapport or winning and retaining the patient's confidence. On the other hand, therapists like Taft (60), Allen (2), and Rogers (44) play up the quality of the counselor-client relationship and are concerned only secondarily with the conversational content aspect of therapeutic interviews. Nonetheless, they are quite insistent that the proper content of counseling contacts is the "feelings" of the patient rather than his overt behavior or his intellectualized beliefs.

What is this content factor in counseling? What are the areas of discussion between counselor and counselee? In line with the foregoing (although at variance with a widespread belief among laymen), therapeutic conversations are concerned with the patient's overt behavior only insofar as it bears on his covert reactions—the anxieties from which he suffers and against which he so non-integratively defends himself.

The client's anxiety (guilt feelings, feelings of inferiority, or inadequacy), then, constitutes the central topic of concern in psychotherapeutic interviews. But clinicians are also interested in the occurrences that engender anxiety. Especially are they interested in the formative past experiences[4] which have been associated with anxiety, and they encourage patients to discuss such events and their reactions to them rather fully. Emphasis throughout seems to be more on the way the client feels about his experience rather than on the objective accuracy of his reportage.

Thus the conversational content of counseling consists chiefly in the discussion of the patient's anxieties and the conditions which either currently evoke them or seem to be causally linked in some historical sense to them.

The relationship aspect of therapeutic procedure has been recently most vigorously expounded by Rogers (44), Snyder (54), and other members (3, 6) of the so-called non-directive or client-centered school. Such a notion is, of course, by no means new to counseling technique. Freud (13) in stressing the idea of transference was talking about essentially the same thing: the basic role in psychotherapy of the affective bonds uniting client to counselor. In the case of orthodox psychoanalysis, transference refers to the displacement of childish attitudes from the analysand's past to the analyst, who becomes a substitute for the important previous objects of his patient's loves and hates. That such things do take place in psychotherapy is not questioned, but whether they *must* occur in just such a form for counseling to be successful may be doubted. For present purposes, it is merely necessary to establish the point that the relationship factor is inherent in the psychoanalytic approach to therapy. Cameron (4), writing from a point of view strongly influenced by Adolph Meyer, says,

---

[4] Even therapists like Rogers, who verbally disclaim any interest in personal history data, hardly prevent their patients' discussing past experiences. It would be revealing to go systematically through a series of electrically recorded non-directive interviews to see if the data collected fall very far short of affording a relatively complete case history. In a preliminary trial by the writer, using material collected from twelve sessions with one case, the greater part of a typical anamnestic form could be filled out from the transcriptions of the recordings.

> . . . . the acquisition of normal biosocial behavior may be greatly facilitated by the organization of a permissive situation, in which the patient has maximal opportunity to work through his attitudes and responses overtly in the presence of a skilled therapist. . . . The immediate goal of treatment in the behavior disorders is that of establishing a biosocial interrelationship . . . in which patient and therapist participate. The ultimate goal is that of making this interrelationship unnecessary and terminating it with benefit to the patient (pp. 576–577).

Dejerine and Gauckler (8) warn, "If . . . you have not been able to awaken a reciprocal sympathy in your patient, and if you have not succeeded in gaining his confidence, it is useless to go any further. The result that you will obtain will be worthless . . . ." Sullivan (58) stresses the concept of parataxis and speaks of the psychiatrist's "participating helpfully in the life of the patient."

While there may be some important differences among the various points of view just touched on, it may be pointed out that there is virtually universal agreement among clinicians on the *importance* of the relationship; there is also high agreement on certain of its characteristics.

The most underscored aspect of the therapeutic relationship seems to be its warmth, permissiveness, and complete freedom from moralistic and judgmental attitudes on the part of the counselor. Far from being a coldly objective consideration of the patient's troubles, therapy necessarily involves a highly personal form of interaction in which the counselor is highly acceptant of the client's behavior, both overt and covert, within clearly defined limits.

Just what "acceptance" means has become somewhat clouded, and a word of clarification may throw some light on the dynamics of the counseling relationship generally. As Sullivan (58) points out, anything a patient feels, says, or does constitutes the data of the therapeutic enterprise. As is the case with data of any kind, one's first job is to understand; it is not to condemn, ignore, reject, or judge. Among such data are the feelings and attitudes that the counselee may develop toward the therapist and which, according to most clinical workers of whatever theoretical orientation, are intimately related to the success or failure of therapy. Here again an atmosphere free from censure or judgment but pervaded by sympathetic understanding is provided by the counselor. On the other hand, acceptance does not imply approval of the client's feelings, attitudes, or overt behavior. This is not surprising since most clinical cases hardly approve of themselves, and their self-disapproval provides one of the most important aspects of the discomfort that brings them into therapy.

As can be inferred from the foregoing, the counseling relationship differs importantly from other forms of human interaction. In the first place, it is essentially one-sided in the sense that the therapist ordinarily says little about himself and that the changes effected within the context of the relationship are centered in the client rather than being a mutual modification. The exchange between counselor and patient, then, does not resemble that between friends in spite of the friendliness that generally permeates the relationship. Secondly, it is sharply limited in that the therapist's expressed interest in his client does not extend beyond the confines of the clinic. The two do not mingle socially, the clinician does not usually intercede for the patient in times of stress, and he generally does not become embroiled in attempts to manipulate the patient's environment. The therapist's office is designated as a place where one can come in perfect safety, free from threats and blame, to "think about" one's problems; but it is not a place where dispensations are sold or intercessions granted. Finally, there is a tacit agreement between therapist and patient that their connection is to be severed as soon as the patient feels free to go about his business without the counselor's support. In other words, the interest, acceptance, and "affection" of the therapist is there for the client to make capital of so so long as he wishes it. Unlike non-clinical situations, there is no pressure on him to maintain the relationship out of politeness or any of the other social rules that more or less govern intimate relationships in society at large.

All this may be recapitualted by saying that the methods common to the various forms of psychotherapy involve (a) the formation of a special kind of personal relationship and (b) a conversation with the patient about his anxieties and the events which tend to produce them. As Finesinger (**10**) puts it, "Communication. . . and the physician-patient relation are the tools that must be adapted to the goals of psychotherapy."

The argument thus far, then, runs something like this: The common problem characterizing clinical patients is anxiety and the behavioral defenses built up against it. The goal of psychotherapy, regardless of the therapist's theoretical leanings, is to eliminate the anxiety and thereby to do away with the symptomatic persistent non-integrative behavior. To accomplish this goal, all therapists use the devices of conversing with the patient about his anxiety and the situations calling it forth both currently and historically, and forming a unique therapeutic relationship. Since all psychotherapies seem to have successes to their credit and since psychotherapy seems to be a process whereby a patient learns to modify his emotional reactions and his overt behavior, it is hypothesized that therapy may be conceptualized from the point of view of general psychology as a problem in learning theory. Such a conceptualization must account for the changes that occur in counselees in terms of these factors that are apparently common to all forms of counseling. Before attempting such a conceptualization, it is necessary briefly to review the situation in learning theory.

MAJOR THEORIES OF LEARNING

One of the major issues with which learning theorists are concerned has to do with the conditions which are necessary if learning is to occur. Two points of view have gained the widest currency with respect to this question.

Reinforcement theory. The first is that of Clark Hull (**19**). Within Hull's system, learning is thought to proceed somewhat in this manner: When a motivated organism is subjected to stimulation—from either or both the stimuli associated with the motivating conditions themselves, as in hunger or pain, and those acting on it from the external environment—it tends to respond in a trial-and-error way. If, in the course of its trial-and-error behavior, the organism performs a response which is associated with the reduction of motivation, the probability of that response's occurring again under similar stimulus conditions is increased, or—to put it somewhat differently—the connection between the present stimuli and the response is strengthened. The central emphasis here is on the occurrence of drive reduction or a satisfying state of affairs, variously designated as the law of effect or the principle of reinforcement. As Miller and Dollard (**31**) succinctly sum it up: To learn, an organism must want something (be motivated in some way), notice something (be acted upon by stimulus cues from the external or internal environment), do something (perform a response or response sequence), and get something (experience a reduction in motivation).

Contiguity theory. Opposed to a reinforcement theory of learning is a point of view which holds that the basic condition necessary for learning is that of contiguity in experience. Tolman and Guthrie are perhaps the outstanding proponents of this theory, although they differ markedly in their conceptions of the nature of learning.

Tolman (**61**), taking his point of departure essentially from *Gestalttheorie*, conceives of learning as the acquisition of information or cognitions about the environment. Variously referred to as "sign-gestalt expectations," "sign-significate relations," and "hypotheses," these cognitions presumably have reference to knowledge which the organism acquires to the effect that a given stimulus or sign, if reacted to in a given way by the organism, will lead to a spatially or temporally more remote stimulus or significate. The necessary condition for the acquisition of such "cognitive maps," as Tolman (62) has called them, is contiguity, the spatial and temporal patterning of stimulus events from sign to significate in the organism's experience. Aided by such secondary principles as recency, emphasis, and belongingness, the law of association by contiguity governs *learning;* learning—i.e., the acquired cognitive maps—together with the organism's needs and skills governs *performance.*

For Guthrie (15) learning is conceived as the acquisition of stimulus-response bonds as is the case with Hull. Unlike Hull, however, he holds that the occurrence of reinforcement is not a necessary condition for learning. Instead, he states that the principle governing learning is association by contiguity: "A stimulus pattern that is acting at the time of a response will, if it recurs, tend to produce that response." Simultaneity of stimulus cues and response is all that is required for the formation of new S-R bonds. Drive states or the existence of unconditioned stimuli are important only as "forcers" of the response to be learned, not as the basis of reinforcement in the Hullian sense.

The behavior with which the various proponents of these points of view have been concerned in their experimentation has consisted for the most part of skeletal muscle acts—maze running, problem-box solutions, conditioned leg flexions, etc. With this fact kept in mind, it seems fair to conclude that the reinforcement point of view seems to have something of an edge in predictive and explanatory utility over contiguity theory. O'Connor (40) has argued rather devastatingly against Guthrie's position by showing that it cannot accommodate the facts of delayed-reward learning. Likewise, Spence and Lippitt (56), Spence and Kendler (55), and Kendler and Mencher (22) have thrown serious doubt on the adequacy of Tolman's notion of contiguity in experience of sign, significate, and response as the essential and sufficient condition for learning.

Reinforcement theory, on the other hand, has demonstrated its utility in a variety of ways. Whiting (65) has conceptualized the socialization process in terms of Hull's notions. Miller and Dollard (31) have made some fruitful incidental remarks on cultural diffusion. Miller (30) has shown the adequacy of the scheme for explaining certain psychopathological phenomena. Loucks (26) and Loucks and Gantt (27) have supplied evidence that strongly supports Hull's contention that the classical conditioning of skeletal muscle responses is merely a special case of learning according to the principle of reinforcement.

It is precisely at this point, however—in the conditioning of defense reactions—that the law of effect runs into difficulties. Hull (18) pointed out this problem as early as 1929, referring to it as "the dilemma of the conditioned defense reaction." He then wrote,

> For a defense reaction to be wholly successful, it should take place so early that the organism will completely escape injury, i.e., the impact of the nocuous (unconditioned) stimulus. But in case the unconditioned stimulus fails to impinge upon the organism, there will be no reinforcement of the conditioned tendency, which means one would expect that experimental extinction will set in at once. This will rapidly render the conditioned reflex impotent, which, in turn, will expose the organism to the original injury. This will initiate a second cycle substantially like the first which will be followed by another and another indefinitely, a series of successful escapes [from all contact with the noxious stimulus] always alternating with a series of injuries. From a biological point of view, the picture emerging from the above theoretical considerations is decidedly not an attractive one.
>
> There is thus presented a kind of biological dilemma . . . (p. 511).

In other words, reinforcement theory finds it hard to explain how an organism can learn to avoid painful stimulation entirely, because if the painful stimuli do not act upon the organism's receptors, no drive is aroused to act as a basis for maintaining the defense reaction.

Mowrer and Lamoreaux (**36**), concerning themselves with this problem, resolved the dilemma by positing a conditioned fear reaction to the conditioned stimulus. According to their formulation, the conditioned stimulus has signal value, signifying to the organism an approaching danger and arousing in it those anticipations of punishment known as the secondary (acquired) drive of fear (anxiety). On the basis of this secondary drive, trial-and-error behavior occurs, out of which is differentiated, according to the principle of reinforcement, a response which reduces the fear and permits the organism to avoid or to minimize the painful unconditioned stimulus.

Such a resolution of the dilemma of the conditioned defense reaction, however, gives rise to another difficulty of comparable magnitude: How is the fear learned? If one holds to a thoroughly monistic reinforcement position, he is forced to say that the drive state of fear or anxiety is somehow "satisfying" or motivation reducing. Baldly, the reinforcement theorist is forced to hold that secondary drive arousal occurs on the basis of drive reduction. That this is certainly contrary to any kind of common sense consideration is immediately apparent, and it is difficult to see how an exchange of one drive for another—the situation which would obtain were the law of effect rigidly adhered to—could be of any biological benefit. This is particularly true when one recalls that many fears, especially neurotic anxiety, are much more debilitating than the objective conditions which generate them—witness the many people who cannot bear to have dental work done or who refuse to see doctors.

Thus, a kind of *impasse* is reached. Reinforcement theory seems to account rather adequately for the acquisition of striped muscle acts; but at least in the conditioned defense situation—most germane to the clinical problems here under scrutiny—its adequacy is dependent on the operation of secondary motivational states, for the acquisition of which it is hard put to it to explain.

Two-factor theory. A number of writers have attempted to overcome this obstacle to efficient theorizing by formulating *two principles* to explain *two different kinds of learning*. Schlosberg (48) in 1937 expressed himself, on the basis of a long series of studies in his laboratory, as believing that there were two types of learning. One had to do with the acquisition of "diffuse, preparatory responses," by which he meant such things as changes in breathing, pulse rate, electrical skin resistance, body volume, voice pitch, and tonicity, which proceeds by "simple conditioning" or according to the principle of association by sheer contiguity. It will be recognized that these reactions are essentially those autonomically mediated viscero-vascular reactions usually thought of as the basic physiological concomitants of emotion. The other type of learning which he felt it necessary to distinguish referred to the acquisition of more "precise, adaptive responses," withdrawal, flexion, or more generally defensive reactions, which are governed by the principle of "success" or reinforcement. These, of course, are the skeletal muscle acts which Hull's kind of theorizing seems to account for so admirably, whether the experimental situation be of the classical or instrumental kind of conditioning.

Skinner (53) in his 1938 volume made explicit a point of view at which he had hinted earlier (52). He distinguished between Type S conditioning as preparatory and Type R as consummatory, holding that the fundamental distinction rested on the event with which the unconditioned stimulus was correlated. In Type S the unconditioned stimulus is correlated with the conditioned stimulus, whereas in Type R it is correlated with the response. Skinner further says,

> Most of the experiments upon skeletal behavior which have been offered as paralleling Pavlov's work are capable of interpretation as discriminated operants of Type R. . . . It is quite possible on the existing evidence that a strict topographical separation of types following the skeletal-autonomic distinction may be made (53, p. 112).

In this formulation, the same classification as that suggested by Schlosberg is implied. Autonomically mediated "emotional" reactions are learned on the basis of contiguity, whereas centrally mediated skeletal muscle responses are learned on the basis of reinforcement.

Razran (**43**) in 1939 offered a somewhat similar formulation, classifying learning according to what he called "quantitative" and "qualitative" conditioning, corresponding to learning without reinforcement and law-of-effect learning. He reports no evidence for the so-called qualitative conditioning of autonomic reactions, but does not say explicitly that quantitative conditioning applies exclusively to the acquisition of viscero-vascular reactions. He does raise the issue of the differential importance of two events, the *application* or onset of the unconditioned stimulus and the *termination* of the unconditioned stimulus, for the conceptualization of types of learning.

More recently, Mowrer (**34**) has vigorously exploited the idea of a two-factor theory of learning to account not only for the learning of skeletal muscle responses but for the acquisition of secondary drives like fear and anxiety. He fully accepts the notion that striped muscle acts, mediated by the central nervous system, are learned, according to the principle of reinforcement, by virtue of their association with the *termination* of the noxious stimulation identified as motivational states. This is not only fully in keeping with Hull's position but is quite in line with Mowrer's own previous enthusiastic experimentation and theorizing as a monistic member of the reinforcement school (**33**). His new point of view, however, holds that smooth muscle and glandular "emotional" reactions, autonomically mediated, are acquired through their association with the *onset* of the paired unconditioned stimulus of pain and conditioned stimulus or signal. In other words, fear refers to the viscero-vascular components of the pain response, conditioned to a substitute stimulus through the latter's contiguity with the *onset* of the action of a noxious adequate stimulus. He prefers to restrict the term *conditioning* to the learning of "emotional" reactions by contiguity and to use *problem-solving* to designate the learning of skeletal responses which "solve" the "problems" created by drives and which are acquired according to the reinforcement principle.

One of the points which must be made immediately with respect to two-factor theories such as these is aimed at the scotching of the criticism often (and fairly) leveled against attempts to account for learning in terms of multiple principles. Such attempts frequently permit the theorist to invoke whichever notion happens most easily to explain his data; he can explain everything but predict nothing. With the possible exception of Razran's, the two-factor formulations just reviewed are not liable to such an attack. While two principles are postulated, contiguity and reinforcement, two learning processes, one involving the viscero-vascular system and the other the skeletal muscular system, are also suggested. The principle that governs one process may not be invoked to explain what occurs in the other. For either process, the theory is monistic and parsimonious and presumably subject to an *experimentem crucis.*[5]

Direct experimental tests of the two-factor theory are as yet few.

One study having an immediate bearing on the issue is that of Mowrer and Suter (**37**). These researchers argue that if the drive-termination theory of acquiring "conditioned" responses is valid, the response should become more readily connected with those stimuli present at the time of drive reduction. If, on the other hand, the drive-onset interpretation is correct, there should be no difference in the resulting learning curves. The rationale on which this deduction is based, of course, is that a conditioned stimulus (warning signal) *must* coincide with or approximate the turning on of the noxious unconditioned stimulus. If this contiguity with the *onset* of drive is all that is necessary for "conditioning" to occur, it should make no difference whether the conditioned stimulus overlaps with the *termination* of the unconditioned stimulus or not. Using an arbitrary running response as an index of fear and as their criterion of conditioning, Mowrer and Suter obtained experimental results confirmatory of their predition: there was no difference in the curves of response acquisition between a group of rats trained under conditions where the conditioned stimulus overlapped and terminated with the turning off of the unconditioned stimulus of shock and a group of animals where the conditioned stimulus was turned off at the time of the unconditioned stimulus's onset.

The interpretation of these results is that the animals learned to *fear* the conditioned stimulus by virtue of its contiguity with the onset of pain. This anticipation of pain gave rise to trial-and-error behavior out of which was differentiated the running response, which was reinforced by fear reduction or the avoidance of pain. The acquisition of the fear reaction was not furthered, as reinforcement theory would predict, by having the warning signal overlap and end in contiguity with the reinforcing state of affairs provided by the termination of the shock.

The more crucial experiment, yet to be done, would involve the testing of the hypothesis that some autonomically mediated reaction, taken as an index of fear, will be attached to some conditioned stimulus by virtue of its association by contiguity with the onset of noxious stimulation, whereas it will not become attached any more effectively under conditions of reinforcement.

Experimentation with viscero-vascular reactions presents many problems, however, and there is little in the literature that can be brought directly to bear on this issue. Indirect evidence is presented in the cited publications of Schlosberg and Skinner and is thoroughly reviewed by Mowrer (**34**).

While such interpretations are not crucial, much recent experimentation on secondary drives is readily assimilable into two-factor theory. Miller (**29**), for example, reports having trained rats by means of strong

---

[5] The two-factor theories reviewed here may be contrasted with those of Stephens (**57**) and Maier and Schnierla (**28**). For a careful and trenchant critique of these points of view, see Kendler and Underwood (**23**).

shock to escape from a white compartment with a grid floor through an open door into a black compartment without a grid. Subsequently, the animals, without shock or noxious stimulation of any kind, learned a new habit—rotating a little wheel to open the door, which had been closed, in order to escape from the white compartment to the black one. This was interpreted to mean that the secondary drive of fear had been acquired and that its termination could be used as reinforcement for striped muscle responses. In terms of the two-factor formulation, the rats learned to run into the black box by virtue of the reinforcement provided by pain reduction. At the same time, however, fear or the visceral component of pain became conditioned to the cues of "whiteness and grid floor" associated with the *onset* of shock. The conditioned fear then served as the drive on the basis of which the wheel-rotating habit was learned without benefit of further primary drive arousal through shock.

It would seem, then, that in spite of its present tentative status, a two-factor theory of learning—holding that adaptive, striped muscle habits are built up according to the principle of reinforcement whereas anticipatory, "emotional" reactions, probably viscero-vascular in nature and having drive properties, are acquired according to the principle of contiguity—has the greatest explanatory and predictive power at the moment.

LEARNING THEORY AND PSYCHOTHERAPY

How can such a conception of learning be applied to psychotherapy to cover the elements of the psychotherapeutic process common to all forms of counseling? It will be recalled that the problem of therapy is essentially that of somehow ridding the patient of neurotic anxiety, which supports his persistent non-integrative defenses and accounts in large measure for his "unhappiness." The tools used by all therapists to accomplish this job are those of conversational content and the therapeutic relationship.

Therapy as the acquisition of symbolic controls. Shaffer (49) suggests that psychotherapy be conceptualized in terms of the patient's acquisition of language symbols by which he can more effectively control his non-integrative behavior. The rationale of this approach is based on the observation than an outstanding characteristic of the maladjusted is their inability to control their own acts; in their own terms, "I know I should (or shouldn't) do this, but I just can't (or must)." Since "normal" people seem to control their behavior by means of symbols—including subvocal and gestural symbols—Shaffer's notion seems at first blush to follow readily.

Such an idea is also more or less explicit in Shaw's (50) analysis of repression and insight. He argues from Mowrer and Ullman's (38) point that,

> The common denominator in all . . . forms of non-integrative behavior seems to be the inability to use symbols appropriately as a means of bringing remote as well as immediate consequences into the present in such a manner that they may exert an influence proportional to their objective importance (p. 81).

Shaw moves from here to the contention that therapy is a process by which non-integrative behavior is eliminated by the making available of symbols, holding that the symbols become cues for the more remote punishing consequences of neurotic defenses.

It is not quite clear, however, according to either Shaffer or Shaw, what the symbolization at which therapy aims might be. If it is the symbolizing of acts which have been repressed, there is no indication of how such a procedure would accomplish anything more than the release of a flood of anxiety heretofore held in check—albeit imperfectly—by the repression mechanism. On the other hand, if the symbols made available by therapy amount only to accurate predictions of the consequences of the client's non-integrative behavior, their utility is questionable on several grounds: First, most clinical patients are only too sharply aware of the self-defeating nature of their activity; their complaint is that they don't know why they engage in it and at the same time seem unable to avoid it. Second, some cases (especially those who have been formally psychoanalyzed) demonstrate a remarkable glibness—sometimes quite accurate—about their own defenses and yet are anxiety ridden on the one hand and socially somewhat obnoxious on the other. It is probably these instances which gave rise to H. M. Johnson's (20) rather oversevere recent strictures on psychoanalysis as therapy and as rationale. Third, there is a question as to whether or not simply making available symbols which can arouse at an earlier point in the temporal sequence the anxiety that accrues from future punishment amounts to anything more than a more effective punishment of the already non-integrative response. In this case, there may be the danger of the repression of one mechanism while another, equally self-defeating, is developed as a defense against a compounded neurotic anxiety, now attached not only to the ineffectively repressed impulses which existed prior to "therapy," but also to those incipient tendencies connected with the defense mechanism which has undergone the "punishment" of having its hurtful ultimate consequences symbolically brought into the psychological present. Thus, if a clinician is dealing with a patient whose anxiety has its origin in the faulty repression of aggressive tendencies and defends himself against it by social withdrawal, the anxiety may be compounded by making the damaging effects of the mechanism more apparent through the providing of symbols within the therapeutic context. All this is not to be construed as an attack on the Shaw-Shaffer hypothesis; as a matter of fact, it seems to describe quite adequately one segment of the therapeutic process. It is merely an effort to point out that such an hypothesis does not seem quite to account for *everything* that happens in psychotherapy.

A somewhat different suggestion, here proposed, is this: If neurotic anxiety is produced by the repression of some unextinguished response, it should follow that the anxiety can be dissipated in one of two ways—either by the elicitation of unreinforced occurrences of the response, thus leading to extinction, or by the connecting of a different affect to response tendencies which have undergone repression. With respect to the illustrative case mentioned above, anxiety could be dispelled either through eliciting self-initiated behavior and failing to reinforce it until extinction occurred, or through forming a bond between the tendencies to self-initiated behavior and some non-anxious visceral reaction which will supplant the connection between anxiety and the repressed activity, In either case, the Shaw-Shaffer notion holds as the first step in therapy, the bringing into communicability (consciousness) of the tendency that has undergone repression.

This lifting of repression is what is usually known as insight. When the patient is able to verbalize the repressed tendencies fundamentally associated with his anxiety, he "sees" or demonstrates insight. It is difficult to understand, however, why this should be equated with cure, regardless of how important it is as a step toward psychological recovery. Merely being able to talk about the cues for anxiety does not make them any less terrifying. Extinction or counter-conditioning is still necessary.

Whether the extinction or the counter-conditioning technique is preferable depends in part on the desirability of the repressed behavior. In the case of self-initiated activity, the question seems rather clear. Socialization has been defined (35) as the process of developing from a dependent infant into an independent and dependable adult. The extinction of tendencies toward self-initiated "responsible" behavior would mean the continuation of dependence and infantilism. It seems probable that few clinicians would look upon this as a suitable therapeutic goal. The same thing might well be said of most of the impulses which typically undergo repression, sexuality being a case in point. The frigid wife, raised under conditions of puritanical restrictiveness, might well find some immediate relief from anxiety by having her repressed sexual impulses extinguished (if this is possible); but it is doubtful that such a procedure would be helpful in her marriage.

The counter-conditioning hypothesis. The hypothesis of counter-conditioning is suggested as somewhat more tenable. It involves the following set of notions: The conversational content aspect of counseling consists in the symbolic reinstatement of the stimuli which produce and have produced the patient's anxiety. Through his words to the therapist, the client, on a symbolic level, again "lives through" the stimulus situations which were painful to him, in which he underwent punishment, and which initiated the repression sequence. This constitutes the lifting of repression, the introduction into communicability of the repressed tendencies, the development of insight. This proceeds essentially by the therapist's reinforcing by his acceptance and his sympathetic participation of the patient's self-revelatory behavior. At the same time, the discussion of the client's anxiety is being carried on within the context of the unique patient-therapist relationship. This is conceived as an unconditioned stimulus for feelings of pleasure, acceptance, security—non-anxious affective reactions. The therapeutic process consists in the establishment of a bond between the symbolically reproduced stimuli which evoke and have evoked anxiety—chiefly the cues associated with the incipient movements toward performing some repressed activity—and the non-anxiety, i.e., comfort and confidence, reactions made to the counseling relationship.

Such a formulation goes somewhat beyond the bounds of "emotional" learning as accounted for by the two-factor theories briefly discussed above. They are chiefly concerned with the learning of fear or anxiety, basic secondary drives. While the idea presented here may be an extension of the theory that its protagonists would find unacceptable, there seems to be no reason why the principle of contiguity should not apply to viscero-vascular reactions that are "pleasant" as well as to those which are "unpleasant"; as a matter of fact, such an application seems to be demanded if the learning of affects is governed by a single principle. The conceptualization proceeds in this wise: Affects possessing drive value—fear, anxiety, and anger[6]—are learned by virtue of the association by contiguity of the visceral aspects of some primary drive with concurrent external stimuli. The so-called "positive" or "pleasurable" affects are learned by virtue of the association by contiguity of proprioceptive cues set up at the onset of drive *reduction* with concurrent external stimuli. It is quite possible that Murray's (**39**) scheme for conceptualizing motivation in terms of goals is analyzable on some such basis as this latter notion.

Hull (**19**) seems to use a similar idea when he defines secondary reinforcement in terms of a stimulus situation which has been closely and consistently associated with the occurrence of need reduction. Experimental animals thus develop "needs" for poker chips, tones of given frequency, black compartments rather than white, etc. Likewise, the judgmental theory of affections, as proposed by Carr (**5**) and expanded upon and experimentally verified by Peters (**41, 42**), is fully consonant with the suggestion here proposed as fundamental in therapy. According to these writers, the pleasantness or unpleasantness of objects is a function of their association with "satisfying" or "unsatisfying" events in experience. Integrating this with the aspect of two-factor theory that deals with the learning of affects, "satisfying" events in experience are those correlated with drive reduction; "unsatisfying" events in experience are those correlated with drive onset.[7]

---

[6] The inclusion of anger in this list of secondary drives is somewhat cavalier. Virtually nothing is known of the conditions under which the learning of anger takes place, and it is certainly not assured that it derives from pain.

To return to the counter-conditioning hypothesis in psychotherapy, a rather striking analogy may be pointed out between this formulation and the now famous experiment of Mary Cover Jones (21) with the boy Peter. It will be remembered that Peter was a three-year-old with a number of acquired fears of various objects, including small white furry animals. In an effort to eliminate these fears, Dr. Jones attempted a counter-conditioning procedure. At lunch time, just as the child began to eat a meal which included his favorite dishes, a white rabbit was introduced in a wire cage at the end of the room, far enough away not to disturb the boy's eating. Each day the animal was brought a little closer until finally Peter could eat with one hand while stroking the rabbit with the other. Further tests showed that the newly conditioned "comfort" reaction to the rabbit had generalized to a large number of other, formerly fear-evoking stimuli such as rats, frogs, cotton, and fur rugs.

The meaning of these results is that a new connection was formed between the stimuli (rabbit) which produced a fear reaction and the comfort reaction made to the stimulus of the lunch with all its various cues. The necessary condition for the formation of this new connection was contiguity of the noxious stimulus and the comfort reaction aroused by the unconditioned luncheon stimulus situation. The problem of how to pair the stimuli so that those connected with the meal did not come to evoke fear does not affect the fundamental point of contiguity as the basis for the establishment of the new bond, but is merely a matter of the spatial and temporal patterning of stimuli common to most experimentation under the conditions of classical conditioning.

The main objection to this analogy probably rests on the point that Peter was troubled by a fear rather than an anxiety—that is, an affective reaction, uncomplicated by repression, made to external stimuli rather than to some impulse to behave in a tabooed way. The objection is certainly granted and actually implies the basis for the first step in therapy, the uncovering by use of the conversational content of therapeutic interviews of the repressed impulses. Before counter-conditioning can occur, the stimuli connected with anxiety must be brought into communicability, where they can be symbolically reinstated at the appropriate times. Insight is a prior condition of counter-conditioning.

A second objection that can be raised to the counter-conditioning notion is this: If therapy is simply a matter of connecting anxiety-provoking stimuli with some comfort reaction, why is it not therapeutically effective to think of one's troubles while lying in a comfortably warm tub?[8] There seem to be three answers to this. First, to a degree it *is* effective. The widespread method of combatting the "blues" by means of a shower is directly in point, as is the use of continuous baths and warm packs in mental hospitals. The real problem is: Why is such a procedure less effective than psychotherapy? This gives rise to the second answer, which is that thinking of one's troubles while lying in a comfortably warm tub is usually of little help in creating insight, symbolically re-introducing the relevant anxiety-producing stimuli. The bath is of little assistance in bringing forbidden impulses into communicability, hence the "therapeutic effects" of the bath are of short duration. The third reply to such an objection is based on the fact that neurotic anxiety is primarily social in its inception. Sullivan (59) insists that this "interpersonal induction of anxiety, and the exclusively interpersonal origin of every instance of its manifestations, is the unique characteristic of anxiety and of the congeries of more complex tensions . . . to which it contributes." This squares perfectly, of course, with the concept of repression and the role it plays in anxiety theory. If neurotic anxiety is an anticipation of punishment for the performance of some tabooed act, it follows that the taboo must have been laid down and enforced through some kind of social medium. Consequently, one would expect in the light of such social origins that the elimination of anxiety would be facilitated by the presence of certain social factors in therapy—provided in this case by the patient-therapist relationship.

---

[7] It is interesting to speculate as to whether or not this is the mechanism underlying the acquisition of aesthetic tastes, preferences, and other "likes" and "dislikes." The implications for a psychological approach to valuative behavior are obvious.

[8] This point was raised in a very helpful personal communication from Dr. John P. Seward. The replies offered to the objection, however, are not chargeable to him.

This last point also bears on the function of catharsis in psychotherapy. It is a commonplace experience among clinicians to have clients say, after a period of vigorous abreaction, "I've thought about that a lot, but I've never said it to anybody before. I feel a bit better now." This poses something of a conceptual difficulty, since it is hard to understand how the expression of an affect should dissipate an affect unless the expression has some effect on the maintaining stimulus conditions. Such an environmental modification certainly does not occur in counseling; and yet catharsis in the social situation of therapy (and possibly in other social situations) seems to bring some relief, whereas catharsis subvocally or made without the presence of a therapist or therapist-surrogate apparently does not. According to the formulation here offered, *catharsis will be effective when it involves (a) the symbolic reinstatement of the repressed cues for anxiety (b) within the context of a warm, permissive, non-judgmental social relationship.* Under these conditions the situation is ripe for counter-conditioning to take place, whereby the patient learns to react non-anxiously to the original stimuli.

The counter-conditioning hypothesis likewise bears on the problems of technique inherent in the directive-non-directive controversy. This argument can perhaps be more profitably stated this way: How much and what can the therapist do to help reinstate symbolically the anxiety-arousing stimuli acting on the patient without endangering the relationship (i.e., weakening the relationship-comfort bond)? Asked in these terms, the question bears on the first step in counseling, that of lifting repressions or developing insight, and becomes the purely empirical matter of determining the categories of counselor response that most effectively further the bringing into communicability of repressed impulses. On somewhat dangerous *a priori* grounds it would seem that interpretation, probing and other more active procedures would be useful unless introduced too preemptorily or too early into therapy, thereby destroying the patient-therapist relationship. That this occurs is not denied, but to attack such techniques as being of no value because they are sometimes misused seems somewhat absurd. The situation is analogous to bringing the rabbit too far into Peter's lunch room too early and connecting the fear reaction to the animal to the stimulus complex of food, room, high chair, and so forth. It seems somewhat nonsensical to argue that the baby should be thrown out with the bath water simply because it is still a bit grimy. One wonders if Peter would have overcome his fear of rabbits had he only been thoroughly "accepted" without ever having any help in reencountering the noxious stimulus in a secure and "pleasant" situation.

The directive-non-directive controversy may well reduce to a consideration of the types of case for which each is best suited. It can be hypothesized that more non-directive approaches will be more likely to succeed with those clients who have few and relatively unsevere repressions, some insight into the sources of their anxiety, and a capacity to relate easily to the therapist. These are cases which do not require much help in *discovering* the anxiety-producing stimuli; they do need assurance from a counselor that they may talk about them in his presence with complete impunity. Conversely, more interpretative methods by hypothesis will be of greater effectiveness with cases characterized by higher defenses, greater repression, and less initial insight. It must be emphasized, however, that all this is a matter of the empirical determination of what techniques work best for given cases so far as the lifting of repressions is concerned. The hypothesis of counter-conditioning is still the means of explaining the diminution of anxiety after insight has been developed.

If this formulation is correct, how can various failures of counter-conditioning methods in psychological treatment be answered? Voegtlin's (63) work with alcoholics is typical. This clinician attempted to cure his patients of drinking by having them take whiskey so heavily dosed with a powerful emetic that vomiting to the point of pain was immediately induced. Results were disappointing. Most of his cases did not build up more than momentary conditioned aversions to alcohol. Of those few who became conditioned against liquor over a period of time, several showed symptom substitutions, e.g., the development of psychosomatic symptoms or neurotic syndromes instead of alcohol addiction.

The first objection to such a procedure is that it consists in a direct attack on the symptomatic mechanism rather than on the underlying anxiety. If the anxiety reduction occurring from drinking were greater than the pain of the treatment, the treatment would have very nearly as little effect as strongly advising the patient "to get on the wagon." The ineffectiveness of "hangovers" is relevant in this connection. Second, if the alcohol addiction were wiped out by virtue of the conditioning procedure, the underlying anxiety would be unaffected, and one would therefore expect that the patient would develop some other persistently non-integrative way of reducing it. Third, the treatment situation contains too many elements of attempting to eliminate a response by merely punishing it. The inefficacy of such methods has already been discussed. Thus, an objection based on such therapeutic experience fails to carry much weight.

Reeducation in psychotherapy. Does the point of view developed here overlook this notion in the therapeutic armamentarium? On the contrary, it fully includes it as an important third aspect of counseling, along with the lifting of repression and the counter-conditioning of anxiety. Following the development of insight, as anxiety is dissipated through conditioning, the patient typically begins to plan. His first tentative steps in this direction may take the form of asking, "What shall I do?" Or it may be a more vigorous exploration of the possible consequences of projected steps. Here the therapist may be of assistance in helping his client to formulate goals clearly and to consider realistically the various behavioral methods he might employ to reach them. This constitutes a law-of-effect learning situation in which reinforcement is produced through the patient's own verbal self-approval or self-disapproval, based in part on the predictions of consequences which the counselor can help him arrive at. In a sense, this constitutes the "rational" exercise of symbolically mediated self-control of which Shaw and Shaffer may be speaking. It is rational insofar as the behavior selected is founded on some consideration of its probable remote outcomes rather than on its immediate value as an anxiety-reducing agent, and it is "responsible" insofar as it is chosen[9] in terms of the patient's own values as of the moment of choice. The counselor does not direct; he merely helps the client work out relatively accurate estimates of the consequences. If a particular behavior pattern is rejected, it merely undergoes a voluntary suppression or is extinguished through failure of reinforcement without being forced into incommunicability and becoming a stimulus for anxiety, as is the case in the repression of punished tendencies. Through this symbolic trial and error, then, the patient develops, according to the principle of reinforcement, a tentative plan of integrative behavior based on rational considerations to supplant his former pattern of persistent non-integrative behavior based on the immediate necessity of reducing anxiety regardless of the ultimate cost.

---

[9] Lest the language used here seem flavored too heavily with free will, reference is made to Hall's (**16**) paper, in which the problem of choice within a deterministic philosophy is discussed.

SUMMARY

A learning theory interpretation of psychotherapy must take into account (a) the fact that all forms of psychotherapy are able to claim cures, (b) the similarity of clinical cases in terms of neurotic anxiety and its defenses, (c) the common goal of psychotherapies of the diminution of anxiety, and (d) the fact that all clinicians employ as their chief techniques conversational content and the therapeutic relationship.

It is here proposed that psychotherapy occurs through three interrelated processes: first, the lifting of repression and development of insight through the symbolic reinstating of the stimuli for anxiety; second, the diminution of anxiety by counter-conditioning through the attachment of the stimuli for anxiety to the comfort reaction made to the therapeutic relationship; and third, the process of reeducation through the therapist's helping the patient to formulate rational goals and behavioral methods for attaining them.

Such a scheme seems to harmonize most effectively with a two-factor learning theory of the type most recently developed by Mowrer (**34**). Such a theory conceives of skeletal muscle responses as being acquired through the principle of reinforcement, whereas viscero-vascular, "emotional" reactions are acquired according to the principle of contiguity.

This formulation is certainly not to be regarded as anything final. It leans rather too much on plausible but inadequately tested hypotheses and on scientifically tenuous analogies. It is offered only as a preliminary attempt to effect a *rapprochement* between psychotherapy and general psychology, and to organize some of the phenomena of clinical practice within the framework of systematic behavior theory.

BIBLIOGRAPHY

1. ALEXANDER, F., & FRENCH, T. *Psychoanalytic therapy.* New York: Ronald Press, 1946.
2. ALLEN, F. *Psychotherapy with children.* New York: W. W. Norton, 1942.
3. AXLINE, V. *Play therapy.* Boston: Houghton Mifflin, 1947.
4. CAMERON, N. *The psychology of the behavior disorders.* Boston: Houghton Mifflin, 1947.
5. CARR, H. *Psychology.* New York: Longmans, Green, 1925.
6. COMBS, A. W. Phenomenological concepts in nondirective therapy. *J. consult. Psychol.*, 1948, **12**, 197–208.
7. DARLEY, J. Review of "*Counseling and psychotherapy.*" *J. abnorm. soc. Psychol.*, 1943, **38**, 199–201.
8. DEJERINE, J., & GAUCKLER, E. *The psychoneuroses and their treatment by psychotherapy.* Philadelphia: J. B. Lippincott, 1913.
9. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57**, No. 3.
10. FINESINGER, J. E. Psychiatric interviewing. *Amer. J. Psychiat.*, 1948, **105**, 187–195.
11. FRENCH, T. Interrelations between psychoanalysis and the experimental work of Pavlov. *Amer. J. Psychiat.*, 1933, **12**, 1165–1203.

12. Freud, S. *New introductory lectures on psychoanalysis.* New York: W. W. Norton, 1933.
13. Freud, S. *A general introduction to psychoanalysis.* New York: Liveright, 1935.
14. Freud, S. *The problem of anxiety.* New York: W. W. Norton, 1936.
15. Guthrie, E. R. A theory of learning in terms of stimulus, response, and association. *National Society for the Study of Education*, 41st Yearbook. Bloomington: Public School Publ. Co., 1942. Pp. 17–60.
16. Hall, E. W. An ethics for today. *Amer. J. econ. Sociol.*, 1943, **2**, 444–446.
17. Horney, K. *The neurotic personality of our time.* New York: W. W. Norton, 1937.
18. Hull, C. A functional interpretation of the conditioned reflex. *Psychol. Rev.*, 1929, **36**, 498–511.
19. Hull, C. *Principles of behavior.* New York: Appleton-Century, 1943.
20. Johnson, H. M. Psychoanalytic therapy versus psychoanalytic rationale. *Amer. Psychologist*, 1948, **3**, 337. (Abstract.)
21. Jones, M. C. A laboratory study of fear: the case of Peter. *Ped. Sem.*, 1924, **31**, 308–315.
22. Kendler, H. H., & Mencher, H. C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, **38**, 82–88.
23. Kendler, H. H., & Underwood, B. J. The role of reward in conditioning theory. *Psychol. Rev.*, 1948, **55**, 209–215.
24. Kraines, S. *Treatment of the neuroses and psychoses.* (2nd ed.) Philadelphia: Lea & Feabiger, 1943.
25. Kubie, L. S. Relation of the conditioned reflex to psychoanalytic technique. *Arch. Neurol. Psychiat.*, 1934, **32**, 1137–1142.
26. Loucks, R. B. The experimental delimitation of neural structures necessary for learning: the attempt to condition striped muscle responses with faradization of the sigmoid gyri. *J. Psychol.*, 1935, **1**, 5–44.
27. Loucks, R. B., & Gantt, W. H. The conditioning of striped muscle responses based on faradic stimulation of dorsal roots and dorsal columns of the spinal cord. *J. comp. Psychol.*, 1938, **25**, 415–426.
28. Maier, N. R. F., & Schnierla, T. C. Mechanisms in conditioning. *Psychol. Rev.*, 1942, **49**, 117–134.
29. Miller, N. E. Studies of fear as an acquirable drive. I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
30. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
31. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
32. Mowrer, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.
33. Mowrer, O. H. The law of effect and ego psychology. *Psychol. Rev.*, 1946, **53**, 321–334.
34. Mowrer, O. H. On the dual nature of learning—A reinterpretation of "conditioning" and "problem-solving." *Harvard educ. Rev.*, 1947, **17**, 102–148.
35. Mowrer, O. H., & Kluckhohn, C. A dynamic theory of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1943. Pp. 69–135.
36. Mowrer, O. H., & Lamoreaux, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, **39**, 29–50.
37. Mowrer, O. H., & Suter, J. Further evidence for a two-factor theory of learning. Unpublished study.

38. MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, **52**, 61–90.
39. MURRAY, H. *Explorations in personality.* New York: Oxford Univ. Press, 1938.
40. O'CONNOR, F. J. Recency or effect? A critical analysis of Guthrie's theory of learning. *Harvard educ. Rev.*, 1946, **16**, 194–206.
41. PETERS, H. N. The judgmental theory of pleasantness and unpleasantness. *Psychol. Rev.*, 1935, **42**, 354–386.
42. PETERS, H. N. Experimental studies of the judgmental theory of feeling: I. Learning of positive and negative reactions as a determinant of affective judgments. *J. exp. Psychol.*, 1938, **23**, 1–25.
43. RAZRAN, G. S. The law of effect or the law of qualitative conditioning? *Psychol. Rev.*, 1939, **46**, 445–463.
44. ROGERS, C. *Counseling and psychotherapy.* Boston: Houghton Mifflin Co., 1942.
45. ROGERS, C. Some observations on the organization of personality. *Amer. Psychologist*, 1947, **2**, 358–368.
46. ROSENZWEIG, S. Some implicit common factors in diverse methods of psychotherapy. *Amer. J. Orthopsychiat.*, 1936, **6**, 412–415.
47. SANFORD, R. N. Psychotherapy and counseling: Introduction. *J. consult. Psychol.*, 1948, **12**, 65–67.
48. SCHLOSBERG, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, **44**, 379–394.
49. SHAFFER, L. The problem of psychotherapy. *Amer. Psychologist*, 1947, **2**, 459–467.
50. SHAW, F. A stimulus-response analysis of repression and insight in psychotherapy. *Psychol. Rev.*, 1946, **53**, 36–42.
51. SHOBEN, E. J., JR. A learning-theory interpretation of psychotheapy. *Harvard educ. Rev.*, 1948, **18**, 129–145.
52. SKINNER, B. F. Two types of conditioned reflex and a pseudo-type. *J. gen. Psychol.*, 1935, **12**, 66–77.
53. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
54. SNYDER, W. U. An investigation of the nature of non-directive counseling. *J. gen. Psychol.*, 1945, **33**, 193–224.
55. SPENCE, K. W., & KENDLER, H. H. The speculations of Leeper with respect to the Iowa tests of the sign-gestalt theory of learning. *J. exp. Psychol.*, 1948, **38**, 106–109.
56. SPENCE, K. W., & LIPPITT, R. An experimental test of the sign-gestalt theory of trial-and-error learning. *J. exp. Psychol.*, 1946, **36**, 491–502.
57. STEPHENS, J. M. Expectancy vs. effect-substitution as a general principle of reinforcement. *Psychol. Rev.*, 1942, **49**, 102–116.
58. SULLIVAN, H. S. *Conceptions of modern psychiatry.* Washington: The William Alanson White Memorial Foundation, 1947.
59. SULLIVAN, H. S. The meaning of anxiety in psychiatry and in life. *Psychiatry*, 1948, **11**, 1–13.
60. TAFT, JESSIE. *The dynamics of therapy.* New York: Macmillan, 1933.
61. TOLMAN, E. C. The determiners of behavior at the choice point. *Psychol. Rev.*, 1938, **45**, 1–41.
62. TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189–208.
63. VOEGTLIN, W. L. The treatment of alcoholism by establishing a conditioned reflex. *Amer. J. med. Sci.*, 1940, **109**, 102.
64. WHITE, R. *The abnormal personality.* New York: Ronald Press, 1948.
65. WHITING, J. W. M. *Becoming a Kwoma.* New Haven: Yale Univ. Press, 1941.
66. WILLIAMSON, E. G. *How to counsel students.* New York: McGraw-Hill, 1939.

Received March 7, 1949.

STATISTICAL METHODS APPLIED TO RORSCHACH SCORES: A REVIEW[1]

LEE J. CRONBACH
Bureau of Research and Service, University of Illinois

While the Rorschach test grew out of clinical investigations, and is still primarily a method of individual diagnosis, there is increasing emphasis on statistical studies of groups of cases. On the whole, the statistical methods employed have been conventional, even though the Rorschach test departs in many ways from usual test methodology. The present review proposes to examine the methods which have been employed to deal with Rorschach data, and to evaluate the adequacy of those often used. It attempts to provide a guide to future investigations by indicating statistically-correct studies which can serve as models. There is no intent here to review the generalizations about the test arising from these studies, or to call into question general research procedures, sampling, and other aspects of the studies.

This report may be considered an extension of a review by Munroe (41). In 1945, she considered the objectivity of previous Rorschach research. She distinguished between the goals attainable by clinical intuitive interpretation and the goals to be reached by more quantitative procedures. She traced the trend in Rorschach literature, noting the gradual decrease in studies based solely on impressionistic treatment of data or on mere counting of scores, and the introduction of significance tests, standard deviations, and other signs of adequate effort to test generalizations statistically. She also pointed out some errors in statistical thinking that lead to faulty conclusions about the Rorschach test. Munroe takes the position, and the writer fully concurs, that statistical research on the Rorschach test is not only justifiable, but indispensable. The flexibility of clinical thinking creates excellent hypotheses, but these hypotheses can only be established as true by controlled studies. Among the propositions suggested by clinical work, some are certainly untrue, due to faulty observation, inadequate sampling, and errors of thinking. Statistical controls are essential to verify theories of test interpretation, and to validate proposed applications of the test. Even though the clinician studying one person makes no use of statistics, he employs generalizations about the test which must rest on scientifically-gathered evidence. Munroe demonstrated that the Rorschach test lends itself to objective studies; the writer reviews the same ma-

---

[1] The writer wishes to express appreciation to Frederick Mosteller and to N. L. Gage, who read this manuscript and contributed many suggestions for its improvement.

terial more technically to evaluate the soundness of the statistical procedures on which the conclusions are based.

CLINICAL TREATMENTS OF DATA

While this paper deals principally with statistical methods applied to *raw* Rorschach data, we shall consider briefly the statistical procedures used when clinically interpreted case records are used in a study. The Rorschach record is usually interpreted qualitatively and in a highly complex manner when the test is given in the clinic, and many studies have been based on these interpreted records. In only a few studies of this type do statistical problems arise.

Dichotomized Rorschach ratings. In one type of study, the interpreter of the records makes a final summary judgment, dividing the records into such groups as "adjusted-maladjusted" or "promising-unpromising," etc. This method is most used for validation studies, where the Rorschach judgment is compared with a criterion of performance or with a judgment from some other test. Simple statistical tests suffice to test the degree of relationship. If the criterion is expressed in two categories (as when the criterion indicates success or failure for each case), chi-square is simple and appropriate. This is exemplified in a study of success of Canadian Army officers (**51**), where a prediction from the Rorschach is compared with a later rating of success and failure. If the criterion is a set of scores on a continuous scale, bi-serial r is usually an adequate procedure. In bi-serial r, one assumes that the dichotomy represents a continuous trait which is normally distributed. This assumption is generally acceptable for personality traits and for ratings of success.

Rorschach ratings on continuous scale. In some studies, the Rorschach interpretation is reported in the form of a rating along a scale, rather than as a dichotomy. When the criterion is dichotomous, bi-serial r is appropriate. (E.g., a prediction of probable pilot success is so correlated with elimination-graduation from training, **21**, p. 632.) For a continuous criterion, like grade-average, product-moment r is conventionally used.

These methods are not entirely satisfactory, because of a limitation of rating scales. If units on the rating scale are not psychologically equal, the correlation may not indicate the full size of the relationship. If ratings are careful, one can assume that men rated "Good" are superior to men rated "Fair," and that men rated "Excellent" are superior to both of these. But it may be unwise to assume that the jump from "Good" to "Excellent" is equal to the jump from "Fair" to "Good," as one automatically does in correlating. One solution to this difficulty is to assume that the trait rated is normally distributed in the men studied. Then we can condense the five-point scale into a dichotomy, which is the case discussed in the preceding paragraph. Alternatively, one may convert the ratings into scaled values which will yield a normal distribution (34). Bi-serial r is then appropriate, if the criterion is dichotomous. Similar reasoning applies to the correlation of a rating with a continuous criterion; one will obtain the most meaningful results by dichotomizing the rating and using bi-serial r, or by normalizing before using product-moment r. These suggestions are summarized in Table I.

TABLE I

PREFERRED METHODS FOR COMPARING RORSCHACH INTERPRETATIONS WITH CRITERIA OF VARIOUS TYPES

Criterion	*Judgment made from Rorschach*	
	Dichotomy	*Continuous scale, unequal units*
Dichotomy	χ^2	χ^2 after dichotomizing rating; r_{bis} after normalizing rating*
Continuous scale, unequal units	χ^2 after dichotomizing criterion; r_{bis} after normalizing criterion*	χ^2 after dichotomizing both variables; r_{bis} after normalizing one, dichotomizing the other; product-moment r after normalizing both
Continuous scale, equal units	r_{bis}*	r_{bis} after dichotomizing rating; product-moment r after normalizing rating

* Point bi-serial must be used if the two parts of the dichotomy cannot reasonably be considered subdivisions of a continuous scale.

Munroe (**42**), comparing a Rorschach adjustment rating with success in academic work, where both variables were reported on a four-category scale, used a coefficient of contingency. Where the correlation surface is nearly normal, this coefficient with proper corrections should give approximately the same result as the product-moment r for normalized data, corrected for broad categories. Yates (**70**) has recently offered an alternative method of adapting the contingency method to take advantage of trends in the relationship between variables expressed as ordered categories.

Matching methods. Another favorite technique for evaluating Rorschach results is blind-matching, which permits a study of each case "as a whole." When a set of Rorschach records (interpreted or not) and another set of data regarding the same individuals are available, one may request judges to match the two sets in pairs. The success of matching is evaluated by a formula developed by Vernon (66). An example of its use is a study by Troup (62), in which judges tried to match two Rorschach records for each person. One hundred fourteen matches were correct out of a possible 120, judges considering five pairs at a time. By the Vernon formula, this corresponds to a contingency coefficient of .88. A coefficient of .40 was obtained when judges attempted to match the record of each case with that of his identical twin. Another excellent illustration of the method is provided by J. I. Krugman (31), who used it to establish that different evaluations of the same Rorschach protocol could be matched, and that the interpretations could be matched to the raw records and to criteria based on a case-study.

The limitations of this method are not statistical; they lie more in the human limitations of judges. A portrait based on the Rorschach may be nearly right, yet be mismatched because of minor false elements. Matching, on the other hand, might be excellent, even perfect; the study would still not guarantee that each element in each portrait was correct, especially if the subjects were quite different from each other. In fact, the portrait might be seriously wrong in some respects, without preventing matching.

A complex modification of the blind-matching method has been proposed and tried by Cronbach (9). Judges are asked to decide whether each statement on a list fits or does not fit a case described in a criterion sketch. Since only about one-third of the statements in the list were actually made about the given case, one can test by chi-square whether the matching is better than chance. The method yields many interesting types of information: (a) an all-over estimate of the validity of predictions with relation to the criterion, (b) a separate estimate of the validity of the description for each case or for subgroups, and (c) an estimate of the validity of statements dealing with any one aspect of personality (e.g., social relations).

ERRORS IN STATISTICAL STUDIES

The majority of statistical studies with the Rorschach test have treated Rorschach scores directly, with clinical judgment eliminated. This is an important type of investigation, which presents numerous problems. Before considering general questions of procedure, however, it is necessary to deal with several errors and unsound practices found in the literature reviewed. These miscellaneous errors must be pointed out lest they be copied by later investigators, and to suggest that the studies in which the errors occurred need to be reevaluated.

Significance tests for small samples. The critical ratio is not entirely satisfactory when applied to small samples. When there are fewer than 30 cases per group, the t test is preferable. This would apply, for example, in Goldfarb's (**19**) comparison of obsessionals with supposedly normal adolescents. His significance ratios are a bit too high, since he used the formula diff./$\sigma_{\text{diff.}}$ with groups of 20 cases. (It may be noted also that Goldfarb's study does not permit sound generalizations about obsessionals as compared to other adolescents. The obsessionals had a mean IQ of 120 compared to 97 for the normals, so that differences between the groups may be due to intelligence rather than obsessional trends.)

Chi-square is generally useful for small samples, but it is important to apply corrections when the number of cases is below 50. This is especially important when the expected frequency in any cell of a 2×2 table is five or lower, under the null hypothesis. Many Rorschach studies fail to recognize the need for corrections, Kaback's (**29**, pp. 24, 38–39) being a striking example. She compares the distribution of such a score as M in each of two groups. To do so, she makes the distribution in a great number of intervals, with only a few cases per interval, and tests the similarity of the distributions by chi-square. In such a case, with many small cell frequencies, no significant result could be expected. Nor is it useful to inquire, as her procedure does, whether the precise distribution of M scores is the same for the two groups (in her case, pharmacists and accountants.) Her major question was whether one group used M more than the other, and this could be answered by dichotomizing the distribution and then applying chi-square, with proper correction. In applying chi-square to the 2×2 tables, one should as a standard practice apply Yates' correction (**56**, p. 169). The importance of this correction will be demonstrated in Table IV. Where groups are dichotomized, it is best to make cuts toward the center, so that marginal totals will remain reasonably large. Special problems in the application of chi-square to successive tests of the same hypothesis, and to problems of goodness of fit, are discussed by Cochran (**6**).

Tests for significance of difference in proportions. Throughout the Rorschach literature, the formula for the significance of differences between proportions is misused. The resulting inaccuracy is slight in most problems, fortunately. This error is common in other work, and even some statistics books appear to endorse the faulty procedure. The usual formula,

$$\sigma_{p_1-p_2} = \sqrt{\frac{pq}{N_1} + \frac{pq}{N_2}}$$

may not be entered with p_1 and p_2, the proportions obtained in the two samples. Instead, one should substitute p_0 for p, where

$$p_0 = \frac{N_1p_1 + N_2p_2}{N_1 + N_2}\,.$$

A significance test inquires whether p_1 and p_2 might arise by chance in sampling from a homogenous population in which the true proportion is p_0 (see **35**, pp. 126–129). Employing p_1 and p_2, instead of entering p_0 in both terms, almost always increases the critical ratio over what it should be. Because no correct model is found in the Rorschach literature, the following example is given using Hertz' data (**25**).

Five boys out of 41, and 0 girls out of 35 gave zero color responses.

$$p_0 = \frac{5+0}{76} = .066$$

$$s.\,d._{\text{diff.}} = \sqrt{\frac{.066 \times .934}{51} + \frac{.066 \times .934}{35}} = .057$$

$$p_1 = .122;\ p_2 = .00;\ \frac{\text{diff.}}{s.d._{\text{diff.}}} = \frac{.122}{.057} = 2.14\ (P = .032)$$

This compares to the critical ratio of 2.41 ($P = .016$) computed by the formula Hertz and other workers have inadvisedly used.

The above computation is equivalent to the determination of significance by chi-square, and yields an identical result. But in this instance the expected frequencies are so low that the correction for continuity becomes important. Applying Yates' correction, we find that P becomes .10, and the reported difference is not significant.

Several studies use the formula for proportions in independent samples when the formula for paired samples should be used. Thus Hertz (**25**), to compare the 12-year-old and 15-year-old records of the same cases, should use a formula for correlated samples as given by Peatman (**44**, p. 407) or by McNemar (**37**; see also **13**, **59**). The correct formula would have yielded significant differences where Hertz found none. Other studies employing matched samples, where the significance of differences was underestimated by a formula for independent groups, are those of Hertzman and Margulies (**27**), Meltzer (**39**), M. Krugman (**32**), Richardson (**48**), and Goldfarb (**20**). In studies where the subjects were children varying widely in age, the proper formula would probably have yielded quite different results.

A study by Brown (**4**) committed this error and one even more serious. He compared records of 22 subjects without morphine and then with morphine. He found that 14 increased in R and 7 decreased. He then treated these as independent proportions of the 22 subjects, computing the critical ratio for the difference 64% minus 32%. These are not proportions in independent samples, and Brown's statistical tests are meaningless. No manipulation of the increase-decrease frequencies is as satisfactory for this problem as the formula given by McNemar. Brown could properly have set a cutting score (e.g., $20R$) and compared the percentage exceeding this level with and without morphine.

Siegel's procedure (**55**), in which the "percentage incidence" of a factor in one group is divided by the incidence in the second group, will be likely to produce misleading results.

An alternative formula for the significance of differences in matched groups is used by Gann (**18**). In applying the formula, however, a serious error was made. The formula given by Engelhart which Gann adopted is

$$\sigma_{\text{diff.}}^2 = (\sigma_{M_1}^2 - \sigma_{M_2}^2)(1 - r_{if}^2)$$

r_{if} is the correlation of the matching variables with the variable in which a difference is being tested. This formula may be extended to differences in proportions, although the estimated population value (p_0) for the proportion should be substituted for M_1 and M_2, as explained above. Gann's major error was to use a value of .9741 for r_{if} in all her calculations. From the context, this seems to be a multiple correlation of all matching variables with *all* dependent variables. The proper procedure, for any single significance test such as the proportion of cases emphasizing W, would be to correlate the matching variables with W-tendency alone. This correlation would almost certainly be close to zero. By the procedure Gann used, the critical ratios are very much larger than they should be. In one comparison where Gann reported a CR of 6.02 the writer has established that the true CR cannot be greater than 2.23, and is almost certainly less.

Comparisons of total number of responses. It is thoroughly unsound to compare the total number of responses of a given type in two samples. Swift (**58**) tested 37 boys and 45 girls. The boys gave a total of 248 F responses; all girls combined gave 246. Swift used chi-square, demonstrating that these 494 responses were divided in a way which departs insignificantly from the theoretical ratio 37:45. But this assumes 494 independent events in her sample whereas she really had 82. The F responses are not independent, since some were made by the same person. She might properly have used the t-test, applied to the means of the groups. The only correct way to use chi-square on her problem is to compare the number of cases exceeding a certain F score (cases, not responses, being the basis of sampling). A similar error has been made by Hertzman (**26**), Rickers (**49**, p. 231), and Werner (**68**).

Richardson (**48**) followed a different erroneous procedure. In her Table 9, she determined what proportion of all responses in each of her groups were W responses, and tested the difference in proportions for significance using the number of subjects in the denominator of the significance formula. The "proportion" she was studying is actually the ratio *Mean* W/*Mean* R, and the standard deviation of this is not correctly given by the formula $\sqrt{pq/N}$. If she must test the W/R ratio, in spite of the difficulties to be considered later, it is necessary to determine the ratio for each person separately and test differences between the groups in one of the conventional ways (e.g., chi-square, t-test, etc.).

Inflation of probabilities. Rorschach studies are peculiarly prone to an error which can arise in any statistical work. If a particular critical ratio or chi-square or t-test corresponds to a P of .05, we conventionally interpret that as statistically significant because "such a value would arise by chance only once in twenty times." While this usually refers to once-in-twenty-samples, it may also be thought of as "once in twenty significance tests," if the several tests are independent. In some Rorschach studies, a vast number of significance tests are computed. Thus Hertz in one study reported the astonishing total of eight hundred significance tests (25). Many of these comparisons reach the one percent level or the five percent level, but even these are not all statistically significant. Quite a few of these differences did arise by chance, and unfortunately we cannot estimate how many because the tests were not experimentally independent. The proper procedure, in such a case, is to recognize that an inflation of P values has taken place. The analogy to monetary inflation is a fair one: The increase in the number of significance tests in circulation causes each P to have less worth than it would normally. We may accordingly raise our "price" arbitrarily, and insist that P reach a higher level than .05 before we label it "significant," and a higher level than .01 before we label it "very significant." Of the differences reported in the Rorschach literature as "significant at the 5% level," probably the majority are due to chance.

There are several ways in which significance levels may be inflated so that they become falsely encouraging. One is the common procedure of testing differences on a great many Rorschach scores. This is of course sound practice, but one must then take the total number of significance tests into account in evaluating P. The inflation is more subtle when the investigator rejects a large number of hypotheses by inspection without computing significance tests, and reports only a few significance tests. Thus Piotrowski and others (46) compared superior and inferior mechanical workers on "all the components used in conventional scoring as well as many others." They finally invented four composite scoring signs on which differences between the two samples were large enough to encourage a significance test. Suppose, for simplicity, that those four tests had yielded P's of .02. The significance of those P's must be minimized in view of the fact that four such differences were found in several hundred implied comparisons which were not actually computed, and two per hundred is chance expectation.

A comparable inflation arises when an investigator slices a distribution in order to take advantage of chance fluctuations and find some "hole" where a test will yield a low P. Hertz applied the formula for significance of the difference in proportions, to compare two groups on $M\%$ (Table II). She introduced a spurious element by slicing the $M\%$ distribution in so many places, and making so many significance tests. If a distribution is dichotomized in many ways, the chances of a "significant" difference rise greatly. Here only one test yielded a P of .05, out of nine attempted. The interpretation "It may be said with certainty, that more girls than boys at 15 years give over 11% M" (25, p. 180) is unjustified. In another sample this fluctuation would not occur. It is not necessary to test explicitly all possible dichotomies for this type of error to arise. If the investigator examines his distribution and makes his cut at the place where the difference is greatest, he has by implication examined and discarded all other possible hypotheses. One of the several studies where this occurs is that of Margulies, discussed later.

Multiple correlation procedures give rise to a similar error. Suppose ten scores are tried as predictors. These scores might be combined in a prediction formula in an infinite number of ways. When an investigator computes correlations and works out the best possible predictive combination for his particular data, he implicitly discards all the other combinations. Even though his combination gives a substantial multiple R for the original sample, it is certain to give a lower correlation in a new sample where the formula can no longer take advantage of chance fluctuations. The common practice of comparing two groups on a large number of signs and developing a checklist score in which the person is allowed one point for every sign on which the two groups differ, is open to the same objection. In a new sample many of these signs will no longer discriminate.[2] When a significance test is applied to a difference in checklist scores or to a multiple correlation in the sample on which the combining formula was derived, the significance test has only negative meaning. If, even after taking advantage of chance differences, one's formula cannot discriminate, it is indeed worthless. But if the result gives a P better than .05, the formula may still be of no value.

TABLE II

SIGNIFICANCE DATA REPORTED BY HERTZ FOR DIFFERENCES IN $M\%$ BETWEEN 15-YEAR-OLD BOYS AND GIRLS (25)

Difference Tested	*Critical Ratio*	*P*
Difference in means	1.47	.15
Difference in medians	2.32	.05
Difference in proportions		
in interval 0–1	.81	..
in interval 0–3	.81	..
in interval 0–5	1.83	.10
in interval 0–7	1.68	.10
in interval 0–9	.90	..
in interval 0–11	2.34	.05
in interval 0–13	1.24	..
in interval 0–15	1.81	.10
in interval 0–17	1.23	..

---

[2] Harris (24) claims that in his experience the Rorschach behaves differently from other tests, and that signs found to differentiate in one sample are usually confirmed in other samples. This appears improbable on logical grounds, and no evidence in the literature supports such a statement.

Rorschach studies which have reported "significant" differences based on an empirical formula without confirming them on fresh samples are those of Montalto (**40**), Harris and Christiansen (**23**), Hertzman, Orlansky, and Seitz (**28**), and Ross and Ross (**52**). Thompson (**60**) reports spurious *r*'s but does not claim significance for them. Buhler and Lefever (**5**, Tables X, XX) mix new cases with the sample used in deriving scoring weights, and therefore fail to provide an adequate test of significance. Significance tests on fresh samples have been properly made by Guilford (**21**), Gustav (**22**), Margulies (**38**), Ross (**50**), and Kurtz (**33**). The latter gives a particularly clear discussion of the issue involved. In most studies, correlations nearly vanish when a Rorschach prediction formula is tried on a new sample.

Still another method of inflating probabilities is to recombine groups of subjects in a way to maximize differences. If one has several types of patients, all of whom earn different mean *M* scores, these groups may be recombined in many ways, and in one of the possible regroupings a pseudo-significant difference may be found. Rapaport and his coworkers (**47**) have carried inflation to bizarre levels. Not only did they consider scores in great profusion and in numerous combinations. They recombined their subjects so that the number of implicit significance tests in their volume is incalculable. They began with subjects in 22 subgroups. Significance tests were then made, on any score, between any pair of subgroups or combinations of them which seemed promising *after* inspection of the data. There were 231 possible pairs of subgroups, and an endless variety of combinations. Thus at times Unclassified Schizophrenics Acute were lumped with one, two, or more of the following: Paranoid Schiz. Acute, Simple Schiz., Uncl. Schiz. Chronic, Par. Schiz. Chr., Uncl. Schiz. Deteriorated; or with all the schizophrenics and preschizophrenics; or with Paranoid Condition, Coarctated Preschiz., Overideational Preschiz., and Obsessive-Compulsive Neurosis. Such willingness to test any hypothesis whatever leaves these workers open to the charge of having regrouped their cases to augment differences. They have undoubtedly reported differences which were created by artificial combinations of chance variations between groups. Every time cases are recombined for a significance test, one must recognize that a large number of implied significance tests were also made, since many other recombinations were rejected without actual computation.

Rorschach studies, because of the great number of scores and the large number of subgroups of subjects involved, are more prone to inflation than other research. The suggestions to be made for sound practice are these:

1. Compare the number of significant differences to the total number of comparisons in the study, both those computed and those rejected by implication.

2. Raise the P value required for significance as the number of comparisons increases.
3. Never accept an empirical composite score or regression formula until its discriminating power has been verified on a new sample.
4. In general, do not trust significance tests unless the hypothesis tested was set up independent of the fluctuations of a particular sample.

These suggestions require that the investigator have clearly in mind the number of comparisons considered. Comparisons are of three types: those rejected as improbable before the data are looked at, i.e., before the study is begun; those not computed because a cursory inspection showed no apparent difference; and those computed. Sometimes the investigator begins with, say, five groups of subjects and ten scores, and frankly wants to unearth all possible differences between types of subjects. Then there are ten ways the groups may be paired against each other, and since each pair may be compared on each score, there are a total of one hundred comparisons in the study. If, on the other hand, the investigator sets out to check only certain relationships—"Schizophrenics differ from neurotics in $F+\%$," "Manics differ from all other groups combined in $FC:CF+C$"—those limited hypotheses may be laid down in advance of the study, and only those comparisons are counted as implied significance tests. To avoid confusion, it is also well for the investigator to specify his cutting point, if a variable is to be dichotomized, before examining the differences between groups. This may be set by an arbitrary rule, for instance that each distribution is to be divided as near to its median as possible, or by an a priori decision to divide at some point such as $2M$. In essence, the investigator must ask himself before he gathers his data, "How many comparisons do I intend to look at, and charge myself for?" A P of .01 may be called significant if it is one of three comparisons charged for, but not if the investigator has looked at three hundred comparisons in order to salvage this one impressive value.

METHODS OF COMPARING GROUPS ON RORSCHACH SCORES

Necessity for Choosing between Statistical Procedures

Because Rorschach scores are numbers which can be added, averaged, distributed, etc., most investigators have used conventional mental-test statistics without question. The most common need for statistics is to compare the test scores of groups and determine the significance of differences. The prominent methods encountered in Rorschach literature are as follows: significance of difference between means (critical ratio or t-test); analysis of variance; bi-serial r; significance of difference in proportions exceeding a particular score, or chi-square; and significance of difference between medians.

Apart from such errors as those listed in the preceding section, there is no reason for considering any of the procedures under discussion as mathematically incorrect. If a significant difference is revealed by any proper significance test, the null hypothesis must be rejected. Nevertheless, the investigator may not choose one of the techniques at random. *Different methods of analyzing the data will lead to different conclusions.* In particular, some procedures lead to a finding of no significant difference even though a true difference could be identified by another attack.

Let us illustrate first with some of Kaback's data (**29**). She administered the group Rorschach to men in certain occupations, and, *inter alia*, compared her groups on the number of popular responses. The mean for accountants is 7.0; for accounting students, 7.3. By the t-test, the difference between means is not significant (P *ca.* .40). (Point bi-serial r applied to the same data gives the same significance level. Point bi-serial r and t are interchangeable procedures, and there is no merit in testing the hypothesis in both ways.) But if she had chosen the chi-square test, quite proper for her data, Kaback would have found a significant difference between the groups. Chi-square would be applied to compare the proportion of cases in each group having five or more popular responses. From her Table IV, this proportion is 60/75, accountants; 72/75, accounting students. The difference between accountants and accounting students is significant ($P <.01$.) In this and other instances, Kaback disregarded a difference when the null hypothesis could be confidently rejected.

Further illustrative data are taken from Hertz' comparison of Rorschach scores of boys and girls. She tested each possible difference by several statistical devices, yielding results such as those for $M\%$ reproduced in Table II. By any of nine methods, she is informed that the two sex groups differ no more than might two chance samples. By the other computations, she is informed that the difference is significant at the 5% level. If different significance tests disagree, what one concludes depends nearly as much on what procedure one adopts as on the data themselves.

Hertz compared her boys and girls in 46 instances. Each time, she tested the significance of differences between means and between medians. Four times the means differed significantly; five times, the medians differed significantly. But in only one out of 46 comparisons was the difference significant by both methods. It is greatly to Hertz' credit that she saw the applicability of more than one significance test. But conclusions of research will be hopelessly confused and contradictory, unless we can find a basis for choosing between the procedures when one says " 'Tis significant" and the other says " 'Taint."

The choice between comparison of means and medians or between the t-test and chi-square cannot be left to the inclination of the experimenter; the whole point of statistical method is to make an analysis freed from subjective judgment. The reason different methods yield different results is that they make different assumptions or try to disclose different aspects of the data. It is therefore important to recognize the ways in which the techniques differ. Differences which are of little concern in connection with most studies have peculiar importance in Rorschach work. The difficulties which make choice of procedures an important problem arise from three causes: the skewness of Rorschach scores, the complications introduced by ratio scores, and the dependence of Rorschach scores on the total number of responses.

Choice of Techniques in View of the Inequality of Units in Rorschach Scales

Many of the significant Rorschach scores give sharply skewed distributions for most populations. This fact is reported repeatedly (**2**, **25**, **47**). Skewness is usually found where many subjects earn 0, 1 or 2 points (i.e., *M*, *FM*, *m*, the shading scores, *CF*, and *C*), and in the location scores *W*, *D*, *Dd*, and *S*. Skewness itself is no bar to conventional significance tests. But in skew distributions the mean and median are not the same. Two distributions may have a significant difference in medians, and not in means (or vice versa) if either is skewed.[3] Furthermore, it is doubtful if a satisfactory estimate of $s.d._{\text{mdn}}$ can be obtained for a skewed distribution.

Disadvantages of the mean and related procedures. In any statistical computation based on addition of scores (mean, *s.d.*, *t*, analysis of variance), numerical distances between scores at different parts of the scale are treated as equal. Thus, since the average of 3 *W* and 7 *W* is the same as that of 1 *W* and 9 *W*, these computations assume that a shift 3 *W* to 1 *W* is equivalent to, or counterbalances, a shift 7 *W* to 9 *W*. There is no way of demonstrating equality of units unless one has some knowledge of the true distribution of the trait in question, or a definition of equality in terms of the characteristics of the property being measured. This problem is present in virtually all psychological tools, but other tests yield normal distributions which are assumed to represent the true spread of ability. On the other hand, *Rorschach interpretation based on clinical experience constantly denies the equality of units for Rorschach scores*. The average *W* score is near 6, and scores from 1 to 10 are usually considered to be within the normal range. No matter how extremely a person is lacking in *W* tendency, his score cannot go below zero. For one who overemphasizes *W*, the score may go up to 20, 30, or more. A *W* score only six points below the mean may be considered clinically to be as extreme in that direction as a score fifteen points from the mean in the other direction. Munroe (**42**) has prepared a checklist which shows how units of certain Rorschach scores would have to be grouped in order to represent a regularly progressing scale of maladjustment. Her groupings based on clinical experience are of approximately this nature:

---

[3] This argument is presented by Richardson (**48**). In attempting to study differences in medians, Richardson unfortunately uses an incorrect method of determining $s.d._{\text{mdn}}$.

W (or *W*%): 0 (or 1 poor) *W* response; 1–14%; 15–60%; 61–100%.
Dd%: 0–9%; 10–24%; 25–49%; 50–100%.
m: 0–1; 2–3; 4–5; 6 or more.

If these units represent increasing degrees of maladjustment, the raw Rorschach scores do not form a scale of psychologically equal units. It is advisable to accept the clinical judgment on this point, especially in the absence of evidence for the assumption of equal units.

Use of median and chi-square. Unlike procedures involving the addition of scores, procedures based on counting of frequencies make no assumption about scale units. In fact, they give the same results no matter how the scale units are stretched or regrouped. The median, or the number of cases falling beyond some critical point (e.g. 10 *W*), depends only on the order of scores. This appears to justify the recommendation that counting procedures such as the median be given preference over additive procedures such as the mean in dealing with skew Rorschach distributions. To test the significance of a difference between two groups, the best procedure is to make a cut at some suitable score, and compare the number of cases in each group falling beyond the cut, using chi-square. This procedure is used by Rapaport (47) and Abel (1). The test of significance of differences between proportions yields the same result (see above). One virtue of cutting scores is that we may test for differences between groups both in the "high" and "low" directions. This is important, since either very high *F*% or very low *F*%, for example, may have diagnostic significance. In the usual analysis based on means, deviations of the two types cancel.

In contrast to the chi-square method, many tests of significance involve computation of the standard deviation. These include the critical ratio of a difference between means or medians, analysis of variance, and the *t*-test. In these procedures, great weight is placed on extreme deviations from the mean. If mean *W* is 6, a case having 25 *W* increases Σd^2 (which enters the computation of the *s.d.*) by about 361 points; a case having 15 *W* increases Σd^2 by about 81 points; and 0 *W*, by only 36 points. In skewed Rorschach distributions, the few cases with many responses in a category have a preponderant weight in determining σ and the significance of the difference. Whether weighting extreme cases heavily is acceptable depends on whether one considers the difference between 15 *W* and 25 *W* to be psychologically large and deserving of more emphasis than, say, the difference from 0 *W* to 5 *W*. Chi-square weights equally all scores below (or above) the cutting point.

Normalizing distributions. One method used to obtain more equal units is to assume that the trait underlying the score is distributed normally in the population. Raw scores are converted to *T*-scores which are normally distributed (**35**, **67**). (This procedure must be distinguished from another conversion, also called a *T*-score, used by Schmidt (**54**). Scores of the type Schmidt used are not normally distributed.) The effect of normalizing is to stretch the scale of scores as if it were made of rubber. Extreme scores below the median are weighted symmetrically to extreme scores above the median. Thus, in the conversion table prepared by Rieger and used by the writer (**10**), the median ($6\frac{1}{2}$ *W*) is placed at 100, and a score of 0 *W* is converted to 66, while 28 *W* becomes 134. This in effect compresses the high end of the *W* scale and expands the low end. This conversion does not alter any conclusion or significance test obtained by dichotomizing raw scores and applying chi-square. But the conversion alters markedly any conclusion based on variance or on comparison of means.

There is obviously much merit in using a procedure which leads to a single invariant result, independent of the assumption of the investigator about the equivalence of scores. Even if scores are normalized it is advised that the median be used to indicate central tendency, and chi-square to test significance. If, for some experimental design, the data must be treated by analysis of variance, the writer believes normalized scores will give results nearer to psychological reality than raw scores, but thus judgment is entirely subjective.

Comparison of mean rank. Attention should be drawn to a new technique invented by Festinger (**14**) which is peculiarly suitable to the problem under discussion. This method assumes nothing about equality of units or normality of distributions, being based solely on the rank-order of individuals. To test whether two groups differ significantly in a score, one pools the two samples and determines the rank of each man in the combined group. The mean rank for each group is computed and the significance of the difference is evaluated by Festinger's tables. The method has not yet been employed in Rorschach research.

The Festinger method and chi-square are not interchangeable. Which should be used depends on the logic of a particular study. Chi-square answers such a question as "Does Group A contain more *deviates* than Group B in the score being studied?" The Festinger method gives weight to differences all along the scale, and therefore asks whether the two groups differ, all scores being considered. In one study, absence of *M* is quite important but differences in the middle of the range have no practical importance. In another study, differences all along the scale are worth equal attention.

The Festinger method appears to have the advantage of greater stability for small samples. Chi-square is much easier to use in samples of 30 or more per group. The Festinger method is not useful when there are numerous ties in score. Further experience with the new method may disclose other important distinctions.

Significance Tests Compared with Estimates of Relationship

Some investigators have perhaps not conveyed the full meaning of their findings to the reader because of a failure to distinguish between tests of the null hypothesis, and estimates of the probable degree of relationship between two variables. The former type of result is a function of the number of cases, whereas the latter is not, save that it becomes more trustworthy as more cases are included. When an investigator applies chi-square, the *t*-test, or the like, he determines whether his observations force him to conclude that there is a relationship between the variables compared. But if the degree of relationship is moderately low, and the number of cases small, the null hypothesis is customarily accepted even though a true relationship exists. It is proper scientific procedure to be cautious, to reject the hypothesis of relationship when the null hypothesis is adequate to account for the data. But in Rorschach studies, where sample size has often been extremely restricted, nonsignificant findings may have been reported in a way which discouraged investigators from pursuing the matter with more cases.

The study of McCandless (36) is a case in point. McCandless compared Rorschach scores with achievement in officer candidate school. In each instance save one, the *t*-test showed *P* greater than .05 that the difference would arise in chance sampling. But the samples compared contained only thirteen men per group. Under these circumstances, it would take a sharply discriminating score to yield a significant difference. If the sample size was raised to about 50 per group, and the differences between groups remained the same, twelve more of McCandless' thirty significance tests would be significant at the five percent, or even the one percent, level. When more cases are added, the differences will certainly change and most of them will be reduced in size. In fact, the writer believes, on the basis of other experience with statistical comparisons of the Rorschach with grades, that McCandless' negative findings are probably close to the results which would be found with a larger sample. But the point is that McCandless, and other investigators using small *N*'s, have submitted the Rorschach to an extremely, perhaps unfairly, rigorous test. One way to compensate for the necessary rigor of proper significance tests is to also report the degree of relationship. A chi-square test may be supplemented by a contingency coefficient or a tetrachoric *r*. A *t*-test may be supplemented by a bi-serial *r*, or point bi-serial (not to determine significance, as Kaback used it, but to express the magnitude of the relationship). Sometimes reporting the means of the groups and their standard deviations, to indicate the degree of overlapping, is an adequate way to demonstrate whether the relationship looks promising enough to warrant further investigation.

To restate the problem: the investigator always implies two things in a comparison of groups: (a) that he considers the null hypothesis is definitely disproven by his data, or else that the null hypothesis is one way to account for the data, and (b) in case the null hypothesis still remains tenable, that he does or does not judge further investigation of the question to be warranted. He can never prove that there is no relationship. So, if his data report a non-significant difference, he must judge whether the difference is "promising" enough to warrant further studies. This judgment is not reducible to rules in the way the significance test is. Whether to recommend further work depends on the difficulty of the study, on the probable usefulness of the results if a low order of relationship were definitely established by further work, and in the investigator's general confidence that the postulated relationship is likely to be found.

Methods of Partialling Out Differences in R

The usual approach when comparing groups is to test the differences in one score after another, and then to generalize that the groups differ in the traits to which the scores allegedly correspond. The various scores, however, are not experimentally independent—a man's total record is obtained at once, and his productivity influences all his scores. If two groups differ in *R*, they may also differ in the same direction in *W* (whole responses), *D* (usual details), and *Dd* (unusual details).

Thus consider the Air Force data in Table III.

TABLE III

RORSCHACH SCORES COMPARED TO SUCCESS IN PILOT TRAINING (**21**, p. 632)

Rorschach Score	*Mean of Successful Cadets*	*Mean of Unsuccessful Cadets*	*Bi-serial r*
R	18.5	15.8	.14
W	9.2	7.3	.24
D	7.1	6.7	.03
W%	60.2	55.8	.08
D%	31.7	37.6	−.15

The first group has more responses than the second. From the means in *W* and *D*, it would appear that the first group has more *W* tendency than the second, but is equal in *D*. But when responsiveness is controlled by converting scores to percentages, the difference in *W* becomes small and the second group is shown to be stronger than the first in emphasis on *D*.

The most striking illustration of this difficulty is Goldfarb's comparison of obsessionals and normals. The obsessional group averages 55 *R*; the normals, 14. Under the circumstances, it is not at all informative to proceed to test *W*, *D*, and *Dd*; all differ significantly in the same direction. One learns nothing about differences between groups in mental approach, which is the purpose of considering these three scores. Most of Goldfarb's other comparisons also merely duplicate the information given by the test in *R*, that is, that the obsessionals are more productive. Although the discrepancy between the groups in *R* is unusually striking in Goldfarb's group, it is present to a lesser but significant degree in a great number of other studies, including those of Buhler and Lefever (5), Hertzman (26), Kaback (29), Margulies (38), and Schmidt (54).

A similar problem complicated Beck's comparison of schizophrenics and normals on *D*. The means were 19.0 and 19.9, respectively; the σ's were 13.5 and 9.9. Beck comments as follows:

> The small difference is accentuated in the very small Diff./S.D. diff: 0.34. There is, however, probably a spurious factor in this small difference. The ogives give us a hint: up to the eighty-second percentile, the curves run parallel, with that for controls where we should expect it, higher. Above this point, the schizophrenics' curve crosses over, and continues higher, and more scattering, as we should expect from the S. D. The spurious element lies undoubtedly in the fact that the schizophrenics' higher response total would necessarily increase the absolute quantity of *D*, since these form the largest proportion of responses in practically all records. Absolute quantity of details is then no indicator of the kind of personality we are dealing with. . . . The medians for *D* are 14.46, 17.2 (2, pp. 31–32).

When one makes several significance tests in which the difference in *R* reappears in various guises, one becomes involved in a maze of seemingly contradictory findings. And interpretation tempts one to violate the rule of parsimony, that an observed difference shall be interpreted by the fewest and simplest adequate hypotheses. To answer the question, how do obsessionals and normals differ? it is simpler to speak of the former as more productive than to discuss three hypotheses, one for each approach factor. And one may certainly criticize Hertzman and Margulies (27) for interpreting differences in *D* and *Dd* between older and younger children as showing the former's greater "cognizance of the ordinary aspects of reality" and greater concern with facts. The older group gives twice as many *R*'s as the former, which is sufficient to account for the remaining differences.

One might argue that *R* is resultant rather than cause, and that the differences in *W*, *D*, *Dd*, etc. are basic. But the Air Force demonstration that *R* varies significantly from examiner to examiner (21) suggests strongly that responsiveness is a partly superficial factor which should be controlled.

Only two studies examine their data explicitly to determine if differences in other categories could be explained in terms of responsiveness alone. Werner (68) found a significant difference in *dd%* between brain-injured and endogenous defectives. But the latter gave significantly more *R*'s. He therefore counted only the first three responses in each card, and arrived at new totals. With *R* thus held about constant, he found the *dd* difference still marked and could validly interpret his result as showing a difference in approach.

Freeman and others (17) found that groups who differed in glucose tolerance also differed significantly in *R*. After testing differences in *M* and *sum C* on the total sample, they discarded cases until the two subsamples were equated in *R*. Since differences between the groups in *M* and *C* were in the same direction even when *R* was held constant, they were able to conclude with greater confidence that glucose tolerance is related to *M* and *C*.

After differences in *R* are tested for significance, it is appropriate to ask what other hypotheses are required to account for differences in the groups. But these other hypotheses should be independent of *R*; otherwise one merely repeats the former significance test and obscures the issue. The usual control method is to divide scores by *R*, testing differences in *W%*, *D%*, *M%*, *A%*, *P%*, etc. Such ratios present serious statistical difficulties discussed in the next section. Moreover, these formulas fail to satisfy the demand for independence from *R*. There may be correlation between *R* and *W%*, etc. (For a sample of 268 superior adults from a study by Audrey Rieger, the writer calculates these *r*'s: $W\% \times R$, −.45, $M\% \times R$, .03, $F\% \times R$, .06. In the latter two cases, there is no functional relation of the percentage with *R*, but the distributions are heteroskedastic. $\sigma_{W\%} = 3.30$ when *R* 5–19 (74 cases) but 2.09 when *R* 40–109 (82 cases). The corresponding sigmas for *M%* are 3.85 and 3.35; for *F%*, 3.23 and 2.29. Only *M%* is really independent of *R*.)

One may control differences in *R* by other methods, provided many cases are available. One procedure is to divide the samples into subgroups within which *R* is nearly uniform (e.g., *R* 20–29), and make significance tests for each such set. A method which requires somewhat fewer cases is to plot the variable against *R* for the total sample or a standard sample, and draw a line fitting the medians of the columns This may be done freehand with no serious error. Then the proportion of the cases in each group falling above the line of medians may be compared by chi-square.

Difficulties in Treating Ratios and Differences

More than any previous test in widespread use, the Rorschach test has employed "scores" which are arithmetic combinations of directly counted scores. One type is the ratio score, or the percentage in which the divisor is a variable score. Examples are *W*:*M*, *M*:*sum C*, *W*/*R*

($W\%$), and F/R ($F\%$). The other type of composite is the difference score, such as $FC-(CF+C)$. In clinical practice, scores of this type are used to draw attention to significant combinations of the original scores; the experienced interpreter thinks of several scores such as FC, CF, and C, at once, placing little weight on the computed ratio or difference. When these scores are used statistically, however, there is no room for the flexible operation of intelligence; the ratios are treated as precise quantities.

It may be noted in passing that a few workers (e.g., **63**) appear to assume that Mean a/Mean b is the same as Mean $\frac{a}{b}$. This is of course not true; the mean of the ratios and the ratio of the means may be quite unequal. One cannot, as Kaback did (**29**, pp. 33, 53, 55), assume that if the ratio of the means is greater for one group than another, the groups differ in the ratio scores themselves. The reader may convince himself by computing the mean ratio for each of the following sets of data in which Mean a/Mean b is constant:

$$\frac{0}{2},\frac{2}{4},\frac{4}{6},\frac{6}{8},\frac{8}{10};\quad \frac{0}{6},\frac{2}{8},\frac{4}{2},\frac{6}{10},\frac{8}{4};\quad \frac{0}{10},\frac{2}{8},\frac{4}{6},\frac{6}{4},\frac{8}{2}.$$

One difficulty with ratio scores is their unreliability. Consider a case with 5 W, 1 M. The ratio $W\!:\!M$ is 5. But M is a fallible score. On a parallel test it might shift to 0 or to 2. If so, the ratio could drop to $2\frac{1}{2}$, or zoom to infinity; such a score is too unstable to deserve precise treatment. The unreliability of another ratio is illustrated in Thornton and Guilford's data (**61**). The reliabilities were, in one sample, .92 for M, .94 for C, but .81 for M/C. In a second sample, the values were .77, .65, and .31. If unreliable ratios are added, squared, and so on, one commits no logical error, but psychologically significant differences become overshadowed by errors of measurement.

Ratios based on small denominators are in general unreliable (**7**). $W\%$ is unreliable for a subject whose R is 12, but relatively reliable for a case whose R is 30. In the former case, addition of one W response raises $W\%$ by 8; in the latter, by 3%. Errors of measurement always reduce the significance of differences by increasing the within-groups variance. A significant difference in $W\%$ might be found for cases where $R>25$. A difference of the same size might not be significant for cases where $R<25$ because of the unreliability of the ratio. If the significance test were based on all cases combined, the difference might be obscured by the unreliability of the ratios in the latter group. One possible procedure is to drop from the computations all cases where the denominator is low. (If there is a significant difference even including the unreliable scores, this need not be done.)

The issue of skewness must again be raised. In the *M*:*sum C* ratio, all cases with excess *C* fall between zero and 1. Those with excess *M* range from 1 to ∞. The latter cases swing the mean and sigma. Following the argument of a preceding section, it is injudicious to employ statistics based on the mean and standard deviation, as McCandless (36) did. By such procedures, different conclusions would often be reached if both *M*: *sum C* and *sum C*:*M* were tested. Procedures leading to a chi-square test are to be recommended, as illustrated in several studies (Rapaport, **47**, pp. 251; Rickers, **49**; etc.) Another solution, less generally suitable, is to convert ratio scores to logarithmic form to obtain a symmetrical distribution (**61**).

A hidden assumption in ratios and differences is that patterns of scores yielding equal ratios (or differences) are psychologically equal. Thus, in $W\%$ the same ratio is yielded by 2 *W* out of 10 *R*, 8 *W* out of 40, and 20 *W* in 100 *R*. One can always define and manipulate any arbitrary pattern of scores without justifying it psychologically, but better conclusions are reached if the assumption of equivalence is defensible. The regression of *W* on *R* is definitely curved. A person with 2 *W* out of 10 *R* is low in *W* tendency, since it is very easy to find two wholes in the cards. Only people with strong tendency and ability to perceive wholes can find 20 *W* in the ten cards, regardless of *R*. As *R* rises above 40, *W* seems to rise very little; the additional responses come principally from *D* and *Dd*. The resulting decline in $W\%$ reflects a drive to quantity, rather than a decreased interest in *W* (cf. **47**, p. 156). Put another way: a strong drive to *W* can easily lead to 90 or 100% *W* when $R<15$; but such a ratio in a very productive person is unheard of. If the regression of a on b is linear and a close approximation to $(a/b) = \textit{some constant}$, ratios may be used as a score with little hesitancy. Otherwise the ratio is a function of the denominator.

This factor is recognized by Munroe, who indicates repeatedly in her checklist that the significance of a particular ratio depends on *R*. Thus 30–40% *M* is rated + if $R=10$, but 16–29% is rated + if $R=50$. Numerically equal Rorschach ratios, then, are not psychologically equal. Rapaport reflects the same point in testing differences between groups in W/D. Instead of applying chi-square to the proportions having the ratio 1:2 or lower, he adjusted his standard.

> In records where *R* is too low or too high, we took cognizance of the fact that it is difficult not to get a few *W*'s and difficult to get too many. Thus, in low

($W\%$), and F/R ($F\%$). The other type of composite is the difference score, such as $FC-(CF+C)$. In clinical practice, scores of this type are used to draw attention to significant combinations of the original scores; the experienced interpreter thinks of several scores such as FC, CF, and C, at once, placing little weight on the computed ratio or difference. When these scores are used statistically, however, there is no room for the flexible operation of intelligence; the ratios are treated as precise quantities.

It may be noted in passing that a few workers (e.g., **63**) appear to assume that Mean a/Mean b is the same as Mean $\frac{a}{b}$. This is of course not true; the mean of the ratios and the ratio of the means may be quite unequal. One cannot, as Kaback did (**29**, pp. 33, 53, 55), assume that if the ratio of the means is greater for one group than another, the groups differ in the ratio scores themselves. The reader may convince himself by computing the mean ratio for each of the following sets of data in which Mean a/Mean b is constant:

$$\frac{0}{2}, \frac{2}{4}, \frac{4}{6}, \frac{6}{8}, \frac{8}{10}; \quad \frac{0}{6}, \frac{2}{8}, \frac{4}{2}, \frac{6}{10}, \frac{8}{4}; \quad \frac{0}{10}, \frac{2}{8}, \frac{4}{6}, \frac{6}{4}, \frac{8}{2}.$$

One difficulty with ratio scores is their unreliability. Consider a case with 5 W, 1 M. The ratio $W\!:\!M$ is 5. But M is a fallible score. On a parallel test it might shift to 0 or to 2. If so, the ratio could drop to $2\frac{1}{2}$, or zoom to infinity; such a score is too unstable to deserve precise treatment. The unreliability of another ratio is illustrated in Thornton and Guilford's data (**61**). The reliabilities were, in one sample, .92 for M, .94 for C, but .81 for M/C. In a second sample, the values were .77, .65, and .31. If unreliable ratios are added, squared, and so on, one commits no logical error, but psychologically significant differences become overshadowed by errors of measurement.

Ratios based on small denominators are in general unreliable (**7**). $W\%$ is unreliable for a subject whose R is 12, but relatively reliable for a case whose R is 30. In the former case, addition of one W response raises $W\%$ by 8; in the latter, by 3%. Errors of measurement always reduce the significance of differences by increasing the within-groups variance. A significant difference in $W\%$ might be found for cases where $R>25$. A difference of the same size might not be significant for cases where $R<25$ because of the unreliability of the ratio. If the significance test were based on all cases combined, the difference might be obscured by the unreliability of the ratios in the latter group. One possible procedure is to drop from the computations all cases where the denominator is low. (If there is a significant difference even including the unreliable scores, this need not be done.)

The issue of skewness must again be raised. In the *M:sum C* ratio, all cases with excess *C* fall between zero and 1. Those with excess *M* range from 1 to ∞. The latter cases swing the mean and sigma. Following the argument of a preceding section, it is injudicious to employ statistics based on the mean and standard deviation, as McCandless (36) did. By such procedures, different conclusions would often be reached if both *M: sum C* and *sum C:M* were tested. Procedures leading to a chi-square test are to be recommended, as illustrated in several studies (Rapaport, **47**, pp. 251; Rickers, **49**; etc.) Another solution, less generally suitable, is to convert ratio scores to logarithmic form to obtain a symmetrical distribution (**61**).

A hidden assumption in ratios and differences is that patterns of scores yielding equal ratios (or differences) are psychologically equal. Thus, in *W%* the same ratio is yielded by 2 *W* out of 10 *R*, 8 *W* out of 40, and 20 *W* in 100 *R*. One can always define and manipulate any arbitrary pattern of scores without justifying it psychologically, but better conclusions are reached if the assumption of equivalence is defensible. The regression of *W* on *R* is definitely curved. A person with 2 *W* out of 10 *R* is low in *W* tendency, since it is very easy to find two wholes in the cards. Only people with strong tendency and ability to perceive wholes can find 20 *W* in the ten cards, regardless of *R*. As *R* rises above 40, *W* seems to rise very little; the additional responses come principally from *D* and *Dd*. The resulting decline in *W%* reflects a drive to quantity, rather than a decreased interest in *W* (cf. **47**, p. 156). Put another way: a strong drive to *W* can easily lead to 90 or 100% *W* when $R<15$; but such a ratio in a very productive person is unheard of. If the regression of *a* on *b* is linear and a close approximation to $(a/b) =$ *some constant*, ratios may be used as a score with little hesitancy. Otherwise the ratio is a function of the denominator.

This factor is recognized by Munroe, who indicates repeatedly in her checklist that the significance of a particular ratio depends on *R*. Thus 30–40% *M* is rated + if $R=10$, but 16–29% is rated + if $R=50$. Numerically equal Rorschach ratios, then, are not psychologically equal. Rapaport reflects the same point in testing differences between groups in *W/D*. Instead of applying chi-square to the proportions having the ratio 1:2 or lower, he adjusted his standard.

> In records where *R* is too low or too high, we took cognizance of the fact that it is difficult not to get a few *W*'s and difficult to get too many. Thus, in low

> R records the 1:2 norm shifted to a "nearly 1:1" while in high R records, the 1:2 norm shifted to a 1:3 ratio (**47**, p. 134).

This adjustment was evidently done on a somewhat subjective basis, and is therefore not the best procedure. It is unfortunate that most other workers have unquestionably assumed that a given score in $W\%$, $M\%$, or $FC-(CF+C)$ has the same meaning regardless of R.

At best, ratio- and difference-scores introduce difficulties due to unreliability and to assumptions of equivalence. There is a fairly adequate alternative which avoids statistical manipulation of ratios entirely. One need only list all significant patterns, and determine the frequency of cases having a given pattern. Thus M: *sum* C can be treated in these categories: coartated (M and C 2 or below); ambiequal, M or $C<2$, M and C differ by 2 or less; introversive, M exceeds C by 3 or more; extratensive, C exceeds M by 3 or more. Any other psychologically reasonable division of cases may be made, and significance of differences tested by chi-square, provided that the hypothesis is not chosen to take advantage of fluctuations in a particular sample. Even this method, however, does not escape the criticism that a given pattern of two scores, such as 3 M, 3 C, has different significance in records where R differs greatly. To cope with this limitation, the pattern tabulation procedure is suggested later.

A detailed consideration of certain work by Margulies is now appropriate, since it affords an illustration of many problems presented above. Her study of the W:M ratio employed a procedure almost like that just recommended, but with departures which are unsound. Margulies compared Rorschach records of adolescents having good and poor school records (**38**). Only her 21 successful boys and her 32 unsuccessful boys need be considered here. She was interested in comparing them on the W:M pattern, in view of Klopfer's belief that this ratio indicates efficient or inefficient use of capacity. She not only tested her data in several ways, but reported the data so that other calculations can be made. Table IV reproduces a part of her data, and shows the results of seven different procedures for determining the significance of the difference.

It should be noted first that Yates' correction is essential for tables with 1 d.f. and low frequencies; in each case where it is applicable, the correction lowers the significance value importantly. Second, attention may be turned to the use of chi-square to test differences between two distributions. Even if more cases were available, it would be unwise to apply chi-square to the distribution cell-by-cell (Procedures 2, 3), since this procedure ignores the regular trend from class-interval to class-interval. Instead, the distribution should be dichotomized. Therefore, procedure 5 is preferable to 2, and 6 is preferable to 3. It will be noted that these recommended procedures indicate higher significance than the tests in which the distributions are compared cell-by-cell.

Margulies is one of the few writers to note the unsoundness of assuming that equal ratios are equal. She pointed out that 20 W: 10 M is not psychologically

TABLE IV

RESULTS OBTAINED WHEN A SET OF DATA IS TREATED BY A VARIETY OF PROCEDURES
(Data from Margulies, **38**, pp. 23, 26, 44)

Distribution I			*Distribution II*			*Distribution III*		
Number of M	*Successful boys*	*Unsuccessful boys*	*W/M ratio*	*Successful boys*	*Unsuccessful boys*	*Pattern of W and M*	*Successful boys*	*Unsuccessful boys*
3 or more	5	5	<1	1	1	$W<6$, M 0–1	0	10
2	9	8	1.00	0	2	$W<6$, $M>1$	8	2
1	3	11	1.1–2.9	8	5	$W>5$, M 0–1	7	9
0	4	8	3.0–4.9	5	7	W 6–10, M 2	3	7
			>4.9	3	9	W 6–10, $M>2$	1	4
			∞ ($W/0$)	4	8	$W>10$, $M>1$	2	0

similar to 2 W: 1 M, and she demonstrated that the regression of M on W is significantly curvilinear. She therefore was properly critical of procedures such as 3 and 6. She next turned to the scatter diagram of M and W, and found successful boys predominating in some regions, and unsuccessful boys in others. After grouping scores into regions as shown in Distribution III, she divided the surface into two areas, one area including cases where W is 0 to 5 and M is 2 or over, plus cases where W is 6 to 10 and M is 3 or over, plus cases where W is over 10 and M is 2 or over. In other words, instead of testing whether the groups are differentiated by a cut along the straight line $M=2$ (Procedure 5), she made her cutting line an irregular one. This hypothesis, tested in Procedure 7, gave apparently quite significant results. The results are of little value, however, since the hypothesis was "cooked up" to fit the irregularities of these specific data. In the cells where W is 6 to 10, and M is 2, there happens to be a concentration of unsuccessful boys. But to draw the cutting line irregularly to sweep in all areas where the unsuccessful predominate is a type of gerrymandering which vitiates a significance test. Hundreds of such irregular lines might be drawn. Therefore, it would be expected that in any sample some line could be found yielding a difference "significant" at the 1% level. At best, the irregular line sets up a hypothesis which, if found to yield a singificant difference in a new and independent sample, could be taken as possibly true.

The law of parsimony enters this problem. Wherever a set of data may be explained equally well by two hypotheses, it is sound practice to accept the simpler hypothesis. Irregular cutting lines, and explanations in terms of patterns of scores, are sometimes justified and neces-

TABLE IV—(*continued*)

Type of analysis	*Procedure*	*Result*	*P*	*Results with Yates' correction* χ^2	*P*
Central tendency	1. Significance of difference in mean M	CR = .70*	.48	..	..
Cell-by-cell comparison	2. Chi-square applied to Distribution I (3 d.f.)	χ^2 = 3.78***	ca. .30	..	..
	3. Chi-square applied to Distribution II (5 d.f.)	χ^2 = 5.30*	ca. .40	..	..
	4. Chi-square applied to Distribution III (5 d.f.)	χ^2 = 17.73*	$<.01$	..	..
Dichotomy	5. Chi-square applied to number of cases with $M>1$ (Dist. I)	χ^2 = 3.46**	.06	2.54**	.11
	6. Chi-square applied to number of cases with $W/M>3$ (Dist. II)	χ^2 = 1.86**	.18	1.13**	.30
Frequency of selected patterns	7. Chi-square applied to frequency having $M>1$ if $W>6$ or >10; having $M>2$ if $6<W<10$ (Dist. III)	χ^2 = 6.58**	.01	5.13**	.03

* Computed by Margulies.

** Computed by the writer.

*** Computed by the writer. Margulies reports 3.64.

sary. But in this case the difference between the groups is explained as well by the hypothesis that the successful boys give more M's as by any non-spurious test of the $W:M$ relationship. Therefore, procedure 5 is the soundest expression of the significance of the Margulies data. With more cases, this difference might be found to be truly significant.

In the above analysis, we find again that different procedures, more than one of which is mathematically sound, give different conclusions. The results from chi-square are less compatible with the null hypothesis than is the critical ratio. Chi-square applied to a dichotomy gives evidence of a possible relationship whereas chi-square applied to the frequency distribution does not. Attention is again drawn to the necessity of regarding with great suspicion any significance test based on a complex hypothesis set up to take advantage of the fluctuations of frequencies in a particular sample. Finally, it is noted that explanations in terms of ratios and patterns should not be sought unless they can account for observed differences more completely than can hypotheses in terms of single scores.

TREATING PATTERNS OF SCORES

Rorschach workers continually stress the importance of considering any score in relation to the unique pattern of scores for the individual. While this is done in clinical practice, there is no practical statistical procedure for studying the infinite complex interrelations of scores and indications on which the clinician relies. Instead of considering the individual patterns, the statistician can at best study certain specific patterns likely to occur in many records. A pattern can be exceedingly complex; there is no statistical reason to prevent one from studying whether (for example) more men than women show high-*S*-on-colored-cards-accompanied-by-emphasis-on-*M*-and-excess-of-*CF*-over-*C*. The only limitation the statistical approach imposes is that the same pattern of scores must be studied in all cases.

Patterns of scores may be considered by means of composite scores, by definition of significant "signs," and by the pattern-tabulation method. The composite score is simply an attempt to express, in a formula, some psychologically important relationship. Examples include the *M*: *sum C* ratio, and the more complex composites developed by Hertz or Rapaport. These scores may be treated statistically like any score on a single category, although most of them are ratios or differences and suffer from the limitations already discussed.

Comparing incidence of "signs." The "signs" approach has been widely used. It is simple and well-adapted to the Rorschach test. Normally, an investigator identifies some characteristic of a special group, such as neurotics, from clinical observation. Then this characteristic is defined in a sign, i.e., a rule for separating those having the characteristic. One such sign, for example, is $FM > M$. After the investigator hypothesizes that some sign is discriminative, the necessity arises for making a test of significance to see if the sign is found more often in the type of person in question. One may soundly compare a new sample of the diagnosed group with a control sample by noting the frequency of the sign in each group and applying chi-square. This procedure is illustrated in studies by Hertzman and Margulies (**38**), and Ross (**50**).

The investigator may invent his own signs, if he follows due precautions to avoid misleading inflation of probabilities. Often it is easier and equally wise to use a predetermined set of signs. The most useful set of signs available at present is the Munroe checklist. She has identified numerous ratios and patterns of scores which she considers significant of disturbance in her subjects (adolescent girls). She has stated that she does not think of her method as a set of signs (**41**), but the difference between her list and others appears to be (a) that it provides an inclusive survey of all deviations in a record and (b) that the list is designed as a whole to minimize duplication from sign to sign. There is no reason why two groups may not be compared by applying the checklist to every record, and then comparing the groups on the frequency with which they receive each of the possible checks. Chi-square is the proper significance test, as used in one of Munroe's studies (**43**). The Munroe signs sometimes are simply defined (e.g. $P-$ is 0 or 1 popular response), but some involve patterns of several scores (thus the sign $FM+$ is defined in terms of FM, M, and R).

Pattern tabulation. Pattern tabulation is a method devised by Cronbach for the study of relations between two or three scores (**10**). It has the advantage of permitting one to study the distribution of patterns in a group. To deal with any set of three scores, e.g. W, D, Dd, one normalizes the three scores for each person, and considers the resulting profile. The profile is expressed numerically in terms of the deviation of the converted scores from their average for each person. These three scores can be plotted on a plane surface, and the resulting scattergram shows the distribution of patterns in a group. If two groups are compared, any type of pattern found more commonly in one group than another can be identified, and the difference in frequency tested by chi-square. The significance level for rejecting the null hypothesis must be set conservatively, as this method involves many implied significance tests. An analysis of variance solution is also possible but not recommended in view of the fact that distributions of patterns are often non-normal.

This method cannot consider hypotheses involving more than three scores at once. It functions best when the three scores are equally reliable and equally intercorrelated. It encounters difficulty due to the fact that some Rorschach scores are unreliable, since any serious error of measurement in one score throws an error into the profile. The method does, however, appear flexible and especially useful for such meaningful patterns as *W-D-Dd* and *M-sum C-F*.

Another group of procedures leading to composite formulas for discriminating groups is treated in the next section.

DISCRIMINATION BY COMPOSITE SCORES

In many problems, it is desired to use the Rorschach to discriminate between two groups. Thus one might seek a scoring formula to predict pilot success, or a "neurotic index" to screen neurotics from a general population. The methods used to arrive at composite scores are the checklist, the multiple regression equation, and the discriminant function.

Checklist scores. The checklist consists of a set of signs. Each person is scored on the checklist and the total number of signs or checks is taken as a composite score. This method has had considerable success, notably in Munroe's study (**42**) and in the formula of Harrower-Erickson and Miale for identifying insecure persons. There are no serious statistical problems in the use of checklists. The total score can be correlated (though eta may be preferable to r). Differences between groups may be tested for significance, preferably by chi-square. Chi-square is advised because a difference in the non-deviate range is rarely psychologically significant; the investigator is usually concerned with the proportion of any group in the deviate range. Buhler and Lefever justifiably apply analysis of variance to their checklist score, to study its ability to differentiate clinical groups (**5**).

Problems do arise, however, in developing checklist scores. A common method is to compare two groups on one raw score after another, noting where their means differ. Each score where a difference arises is then listed as a sign, and counted positively or negatively in obtaining the checklist score for each case. This method takes advantage of whatever differences between samples arise just from accidents of sampling. If sample A exceeds B in mean *M*, allowing one point in the total score for high *M* will help discriminate A's and B's. In this sample, the A's will tend to earn higher checklist scores. But often in a new sample such a difference will not be confirmed, and the *M* entry in the composite will not discriminate.

One study employing the sign approach should be pointed out to Rorschach workers. Davidson (**12**) sought to determine the relationship between economic background and Rorschach performance in a group of highly intelligent children. Her treatment of data is noteworthy because of the flexibility of her procedures; statistics are applied with great intelligence, new procedures being adopted for each new type of comparison. While the reviewer disagrees with some of the judgments she made in selecting procedures, her treatment is free from overt errors and well worth study by other Rorschach investigators.

Davidson divided her 102 cases among seven economic levels. She studied the Rorschach performance in various ways. First, she made a clinical analysis of each child, and placed him in one of nine categories (introvert adjusted, childish, constricted, disturbed, etc.). The distribution which resulted is a 7×9 table. Recognizing that the expected frequency in each cell is quite small, she combined groups to form a 3×3 table before applying the chi-square test for significance. This same type of condensation would have been advisable in some other comparisons she made, such as that between personality pattern and IQ.

Davidson next applied a list of signs, and obtained for each case the total number of signs of maladjustment. The number of signs was correlated with economic level, and the correlation was shown not to differ significantly from zero. She tested the significance of the difference in mean number of signs by the critical ratio. These procedures appear well suited to her data. A third attack on the data treats one Rorschach score at a time. Here Davidson placed her cases in seven categories, ranging from highest to lowest economic level. By analysis of variance, she demonstrated that differences among the seven groups were significant only for a few of the scores. The application of analysis of variance to continuous data appears to have been an unwise decision. Analysis of variance, like chi-square or eta applied to a variable divided in several categories, ignores the order of the categories. Consider the following set of means in the score *M – sum C*:

Economic level	1	2	3	4	5	6	7	Total
Mean score	1.17	1.86	1.29	0.96	−0.75	−0.13	−0.71	0.63

The downward trend from Group 1 to Group 7 gives great support to the hypothesis that this score is related to economic level. Analysis of variance estimates significance without considering this trend; the same significance estimate would be arrived at if Group 2 had had the mean of −0.13 and Group 6 the mean of 1.86. Davidson might have computed the correlation between each score and the economic level, but the skewness of some Rorschach scores weighs against this suggestion. The simplest procedure for testing this trend is to split the group into a 2×2 table by combining adjoining categories in the economic scale, and dichotomizing the Rorschach score at a convenient point. Chi-square would then give the significance estimate. Such a procedure might have yielded significant differences in several instances where Davidson found none.

In justice to Davidson, it should be repeated that her data have been singled out for critical comment because of their exactness and completeness, rather than because they were improperly handled. The foregoing suggestions point to ways in which she might have arrived at additional important findings.

The multiple regression formula. A limitation of check lists is that they are simple additive combinations of signs which individually discriminate. But in such a composite a given trait may enter several times if it is reflected in several signs, and thus have greater proportionate weight then it deserves. The checklist method does not allow for the possibility that certain signs may reinforce each other to indicate more severe maladjustment than is indicated by a combination of two other non-reinforcing signs, or for the possibility that two signs which are individually unfavorable may operate to neutralize each other. Multiple regression and the discriminant function are more powerful procedures than the usual checklist score, because they consider the intercorrelations of scores and weight them accordingly.

By multiple correlation, one arrives at a regression equation which assigns weights to those variables which are correlated with a criterion and relatively uncorrelated with each other. This formula may be used to predict or to discriminate between groups. One such formula is that of the Air Force, used in its attempt to predict pilot success (**21**):

$$2(Dd+S\%)+6\ FM+8\ W-1.5\ D\%+R-(VIII-X\%).$$

Multiple correlation does not seem especially promising for Rorschach studies. Even such an elaborate formula as that above turns out to have little or no predictive value when applied to a fresh sample. Even if it were stable, any formula of this type must assume that strength in one component compensates linearly for weakness in another. In this formula, emphasis on *Dd* would cancel weakness in *FM*, in estimating a man's pilot aptitude. It is most unlikely that the factors cancel each other in the personality itself. The simple linear regression formula provides an efficient weighting if the assumption of linear compensation is valid, but interrelations between aspects of personality are probably far too complex to be adequately represented in this way. The most that can be said for a regression formula is that, when derived on large samples (and this may require 5000 cases), it is a more precise prediction formula than the simple checklist score can be. It cannot hope to yield very accurate predictions if interrelations within personality are as complex as Rorschach interpreters claim.

The discriminant function is a relatively new technique giving a formula which will separate two categories of men as thoroughly as possible from a mixed sample. It would be used to develop an effective index for separating good from poor pilots (not for predicting which man will be best, as the regression formula does) or for distinguishing organics and feeble-minded. A practical procedure for dealing with multiple scores has just been published by Penrose (**45**), and has not been employed in Rorschach research. It appears likely to have real value in studies comparing different types of subjects.

Like the regression formula, however, the discriminant function provides a set formula. In this formula, it is assumed that one factor compensates for or reinforces weakness in another factor. The interactions within personality are probably too complex to be fully expressed by linear or quadratic discriminant functions.

CORRELATION AND RELIABILITY

Correlations of scores. For one purpose or another several studies have tried to show the relationship between the several Rorschach scores or between Rorschach scores and external variables. The conventional procedure for showing that two characteristics are associated is to compute a product-moment correlation between the variables. This has been done by Kaback (**29**), Vaughn and Krug (**64**), and others.

This method is unable to show the full relationship between variables when the regression of one on the other is curvilinear. Such a regression often occurs when one variable or both have a sharply skewed distribution. In fact, Vaughn and Krug note that one of their plots is curvilinear. The extent to which association may be underestimated is suggested by the following data. The data used are taken from tests administered individually by Audrey Rieger to several hundred applicants for employment, usually for managerial or technical positions. The tests were carefully scored by the Beck method. Generalization from the data must be limited because the group is not a sample of any clearly defined population. For 268 men, the product-moment correlation between *D* and *Dd* is .735. The curvilinear correlations are η_{DdD}, .785; η_{DDd}, .823. There is significant curvilinearity. If *D* and *Dd* are normalized, the regression becomes linear except for the effect of tied scores where $Dd=0$; for the converted scores, $r=.767$.

Brower employs rank-difference correlations in comparing certain Rorschach scores to physiological measures (3). This is a useful method for small samples and is equally sound for linear and non-linear regressions. Thus, a rank-correlation of *W/M* with another score is the same except for sign as the correlation for the inverted ratio *M/W*, but the product-moment correlations are far different.

The rank method does have the disadvantage of weighting heavily the small and unreliable differences in the shorter end of skew distributions, where many cases have the same rank. This might lower the correlations for a score like *Fc*, but is not a difficulty with scores distributed more symmetrically over a wide range, such as *F* or *VIII–X%*. Normalizing has the same disadvantage. This is a reflection of the inability of the test to discriminate finely among cases in the modal end of a severely skewed distribution.

Reliability coefficients. Test reliability is ordinarily estimated by the retest or the split-half method. These methods are not very appropriate for the Rorschach test, the former because of memory from trial to trial, the latter because the test cannot be split into similar halves. Nevertheless, both methods have been used in the absence of better procedures.

The split-half method introduces a statistical problem which not all investigators have noted, namely, that the Spearman-Brown formula must not be applied to ratios with variable denominators such as *W%* and *M/sum C*. Methods for estimating the reliability of ratio scores have been treated elsewhere (7, 8), but these procedures are not useful when the denominator is relatively unreliable (as in *M/sum C*).

It is desirable to estimate reliability of scores separately for records of varying length. Vernon (65) found that Rorschach scores were much more reliable for cases where $R>30$ than when $R<30$. This implies that it is unsatisfactory to estimate just one reliability coefficient for a group with varied R. Instead, the standard error of measurement of W, or $W\%$, should be determined separately for cases where $R=10$–15, $R=15$–25, $R=25$–35, or some such grouping.

The reliability of patterns of scores is a difficult problem. If both M and W were perfectly reliable, any pattern or combination based on the two scores would also be perfectly reliable. But these scores are unstable; subjects vary from trial to trial in M or W or both. Nevertheless, Rorschach users insist that the "pattern" of scores is stable. If there is any substance to this claim, it means that certain definable configurations of the scores are stable even though the separate scores are not. The configurations may be as simple as the W/M ratio or may be complex structures of several scores. One may establish the reliability of any composite score by obtaining two separate estimates from independent trials of the test.

The method of determining reliability by independent estimates has rarely been used. A study by Kelley, Margulies, and Barrera (**30**) is of interest, even though based on only twelve cases. The Rorschach was given twice, and between the trials a single electroshock was given, reportedly sufficient to wipe out memory of the first trial without altering the personality. In the records so obtained, R shifted as much as 50 per cent from trial to trial, and absolute values of some other scores shifted also. In several cases where scores shifted, it can be argued that the *relationship* between the scores did not shift and that the two records would lead to similar diagnoses. The authors made no attempt at statistical treatment. Probably this ingenious procedure will rarely be repeated. Useful studies could certainly be made, however, by comparing performance on two sets of inkblots without shock (cf. Swift, **57**). Even if the two sets are not strictly equivalent, the data would indicate more about the stability of performance than any methods so far employed.

At first glance, it appears logical to set up composite scores, obtain two separate estimates, and correlate them. Even this is unsuitable for Rorschach problems, however. As pointed out before, a given ratio such as 20% W or W/M 2.0 has different meaning in different records, depending on the absolute value of W. The pattern might conceivably be defined by a curvilinear equation, but this becomes unmanageable, especially as several variables enter a single pattern. The problem is one of defining when two patterns are psychologically similar, and of defining the magnitude of the difference when they are not equivalent. No one would contend that the W/M balance is unchanged if a subject shifts from 12 W: 2M to 60 W: 10 M. The problem is to define and measure the balance in a numerical way. The approach pattern W-D-Dd has three dimensions. If we wish to estimate reliability by comparing two sets of these three scores we have a six-dimensional array, for which no present methods are adequate. So far, even the pattern-tabulation method reduces such data only to four dimensions, which leaves the problem still unmanageable. All that can be recommended is that additional attention be given to this challenging problem. We can now obtain adequate evidence on the stability of Rorschach patterns only by such a method as Troup's (**62**), discussed in the first section of this paper. It will be recalled that she had two sets of records interpreted clinically, and employed blind-matching to show that the inferences from the Rorschach remained stable.

Two unique but entirely unsound studies by Fosberg (**15**, **16**) employed a novel procedure to estimate the reliability of the total pattern. He gave the test four times, under varied directions. He then compared the four records for each person. In one study he used chi-square to show that the psychograms for each person corresponded. But this statistical test merely showed that the D score in record 1 is nearer to D in record 2 than it is to W, C, or other scores. That is, he showed that the scores were not paired at random. But, since each score has a relatively limited range for all people—i.e., D tends to be large, m tends to be small, etc.—he would have also obtained a significantly large chi-square if he had applied the same procedure to four records from *different* persons. One may also point out that finding a P of .90 does not prove that two records do come from the same person, but only that the null hypothesis is tenable, or possibly true. Fosberg's second study, using correlation technique, is no sounder than the first. Here the two sets of scores for one person were correlated. That is, pairs of values such as W_1-W_2, D_1-D_2, etc. were entered in the same correlation chart. As before, the generally greater magnitude of D causes the two sets to correlate, but high correlations would have been obtained if the scores correlated came from two different subjects.

Objection must also be made to several procedures and inferences of Buhler and Lefever (**5**), in their attempts to demonstrate the dependability of their proposed Basic Rorschach Score. (1) They used the split-half method on the total score, by placing half the signs in one list, the other half in a second list, and scoring each person on both lists (**5**, p. 112). They then correlated the two halves to indicate reliability. Because the correlation was computed on cases used to determine the scoring weights for the items, the resulting correlation is spuriously high. Even if new cases were obtained, the split-half method would be incorrect because the checklist items are not experimentally independent. A single type of performance enters into a great number of separately scored signs (in their checklist, M affects items 1, 2, 5, 6, 7, 8, 10, 11, 12, 51, 52, 53, 86, 93, 94, 95, 96, 99, 100, 101, and 102). A "chance" variation in M would alter the score on all these categories, and would spuriously raise the correlation unless these linked categories were concentrated in the same half of the test. (2) They derived separate sets of weights from the comparison of Normals vs. Schizophrenics, Nurses vs. Schizophrenics ,and other groups. The correlation between the scoring weights is high, which they take as evidence for reliability (pp. 112 ff.). At least one serious objection is that the weights were derived in part from the same cases. If, by sampling alone, FK happened to be rare among the Schizophrenic group, this would cause the sign FK to have a weight in both the Normal-Schizophrenic key and the Nurse-Schizophrenic key. The evidence is not adequate to show that the weights would be the same if the two keys were independently derived. This objection does not apply to another comparison of the same general type, where the four samples involved had no overlap. (3) Certain papers were scored repeatedly, using sets of weights derived in comparable but slightly different ways (p. 116). The correlations of the resulting sets of scores are advanced as evidence of reliability. Any correlation of separate scorings of the same set of responses is in part spurious. If responses of individual subjects were determined solely by chance, there would still be a correlation when keys having any similarity to each other were applied to the papers. The reliability of the performance of the subject, and that is what reliability coefficients are supposed to report, cannot be revealed by rescorings of the same performance.

CONCLUSIONS

The foregoing analysis and the appended bibliography are convincing evidence that Rorschach workers have sought statistical confirmation for their hypotheses. But the analysis also shows that the studies have been open to errors of two types: (1) erroneous procedures have led to claims of significance and interpretations which were unwarranted; and (2) failure to apply the most incisive statistical tests has led workers to reject significant relationships. So widespread are errors and unhappy choices of statistical procedures that few of the conclusions from statistical studies of the Rorschach test can be trusted. A few workers have been consistently sound in their statistical approach. But some of the most extensive studies and some of the most widely cited are riddled with fallacy. If these studies are to form part of the base for psychological science, the data must be reinterpreted. Perhaps ninety per cent of the conclusions so far published as a result of statistical Rorschach studies are unsubstantiated—not necessarily false, but based on unsound analysis.

Few of the errors were obvious violations of statistical rules. The Rorschach test is unlike conventional instruments and introduces problems not ordinarily encountered. Moreover, statistical methods for such tests have not been fully developed (11). It is most important that research-workers using the Rorschach secure the best possible statistical guidance, and that editors and readers scrutinize studies of the test with great care. But statisticians have a responsibility too, to examine the logic of Rorschach research and the peculiar character of clinical tests, in order to sense the limitations of conventional and mathematically sound procedures.

Present statistical tools are imperfect. And no procedure is equally advisable for all studies. Within these limitations, this review has suggested the following guides to future practice.

1. Matching procedures in which a clinical synthesis of each Rorschach record is compared with a criterion are especially appropriate.

2. If ratings are to be treated statistically, it is often advisable to dichotomize the rating and apply chi-square or bi-serial r.

3. Common errors which must be avoided in significance tests are:
   a. Use of critical ratio and uncorrected chi-square for unsuitably small samples.
   b. Use of sample values in the formula for differences between proportions.
   c. Use of formulas for independent samples when matched samples are compared.
   d. Interpretation of P-values without regard for the inflation of probabilities when hundreds of significance tests are made or implicitly discarded.
   e. Acceptance of conclusions when a significant difference is found with a hypothesis based on fluctuations in a particular sample.

4. Counting procedures are in general preferable to additive methods for Rorschach data. The most widely useful procedures are chi-square and analysis of differences in mean rank. These yield results which are invariant when scores are transformed.

5. Normalizing scores is frequently desirable before making significance tests involving variance.

6. Where groups differ in total number of responses, this factor must be held constant before other differences can be soundly interpreted. Three devices for doing this are: rescoring a fixed number of responses on all papers, constructing subgroups equated on the number of responses, and analyzing profiles of normalized scores (pattern tabulation).

7. Ratio and difference scores should rarely be used as a basis for statistical analysis. Instead, patterns should be defined and statistical comparisons made of the frequency of a certain pattern in each group. Use of chi-square with frequencies of Rorschach "signs" is recommended.

8. Multiple regression and linear discriminant functions are unlikely to reveal the relationships of Rorschach scores with other variables, since the assumption of linear compensation is contrary to the test theory.

9. Rank correlation, curvilinear correlation, or correlation of normalized scores are often more suitable than product-moment correlation.

10. No entirely suitable method for estimating Rorschach reliability now exists. Studies in this area are much needed.

There are in the Rorschach literature numerous encouraging bits of evidence. The question whether the test has any merit seems adequately answered in the affirmative by studies like those of Troup, Judith Krugman, Williams (**69**), and Munroe. Supplemented as these are by the testimony of intelligent clinical users of the test, there is every reason to treat the test with respect. One cannot attack the test merely because most Rorschach hypotheses are still in a pre-research stage. Some of the studies which failed to find relationships might have supported Rorschach theory if the analysis had been more perfect. How accurate the test is, how particular combinations of scores are to be interpreted, and how to use Rorschach data in making predictions about groups are problems worth considerable effort. With improvements in projective tests, in personality theory, and in the statistical procedures for verifying that theory, we can look forward to impressive dividends.

BIBLIOGRAPHY

1. Abel, T. M. Group Rorschach testing in a vocational high school. *Rorschach Res. Exch.*, 1945, **9**, 178–188.
2. Beck, S. J. Personality structure in schizophrenia. *Nerv. and ment. Dis. Monogr.*, 1938, No. 63.
3. Brower, D. The relation between certain Rorschach factors and cardiovascular activity before and after visuo-motor conflict. *J. gen. Psychol.*, 1947, **37**, 93–95.
4. Brown, R. R. The effect of morphine upon the Rorschach pattern in post-addicts. *Amer. J. Orthopsychiat.*, 1943, **13**, 339–342.
5. Buhler, C., Buhler, K., & Lefever, D. W. *Rorschach standardization studies. Number I. Development of the Basic Rorschach Score.* Los Angeles: C. Buhler, 1948.
6. Cochran, W. G. The chi-square correction for continuity. *Iowa St. Col. J. Sci.*, 1942, **16**, 421–436.
7. Cronbach, L. J. The reliability of ratio scores. *Educ. psychol. Msmt.*, 1941, **1**, 269–278.
8. Cronbach, L. J. Note on the reliability of ratio scores. *Educ. psychol. Msmt.*, 1943, **3**, 67–70.
9. Cronbach, L. J. A validation design for personality study. *J. consult. Psychol.*, 1948, **12**, 365–374.
10. Cronbach, L. J. Pattern tabulation: a statistical method for treatment of limited patterns of scores, with particular reference to the Rorschach test. *Educ. psychol. Msmt.*, in press.
11. Cronbach, L. J. Statistical methods for multi-score tests. Paper presented before the Biometrics Section, American Statistical Association, December, 1948. To be published.
12. Davidson, Helen H. *Personality and economic background.* New York: King's Crown Press, 1945.
13. Edwards, A. L. Note on the "correction for continuity" in testing the significance of the difference between correlated proportions. *Psychometrika*, 1948, **13**, 185–187.
14. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 97–105.
15. Fosberg, I. A. Rorschach reactions under varied instructions. *Rorschach Res. Exch.*, 1938, **3**, 12–31.
16. Fosberg, I. A. An experimental study of the reliability of the Rorschach technique. *Rorschach Res. Exch.*, 1941, **5**, 72–84.
17. Freeman, H., Rodnick, E. H., Shakow, D., & LeBeaux, T. The carbohydrate tolerance of mentally disturbed soldiers. *Psychosom. Med.*, 1944, **6**, 311–317.
18. Gann, E. *Reading difficulty and personality organization.* New York: King's Crown Press, 1945.
19. Goldfarb, W. A. A definition and validation of obsessional trends in the Rorschach examination of adolescents. *Rorschach Res. Exch.*, 1943, **7**, 81–108.
20. Goldfarb, W. Effects of early institutional care on adolescent personality. *Amer. J. Orthopsychiat.*, 1944, **14**, 441–447.
21. Guilford, J. P. (Ed.) *Printed classification tests.* AAF Aviation Psychology Program Research Reports, No. 3. Washington: Government Printing Office, 1947.
22. Gustav, Alice. Estimation of Rorschach scoring categories by means

of an objective inventory. *J. Psychol.*, 1946, **22**, 253–260.

23. Harris, R. E., & Christiansen, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, **21**, 269–284.
24. Harris, T. M. The use of projective techniques in industrial selection. In *Exploring individual differences*, American Council on Education Studies, Series 1, No. 32, 1948. Pp. 43–51.
25. Hertz, Marguerite R. Personality patterns in adolescence as portrayed by the Rorschach ink-blot method: I. The movement factors. *J. gen. Psychol.*, 1942, **27**, 119–188.
26. Hertzman, M. A comparison of the individual and group Rorschach tests. *Rorschach Res. Exch.*, 1942, **6**, 89–108.
27. Hertzman, M., & Margulies, H. Developmental changes as reflected in Rorschach test responses. *J. genet. Psychol.*, 1943, **62**, 189–215.
28. Hertzman, M., Orlansky, J., & Seitz, C. P. Personality organization and anoxia tolerance. *Psychosom. Med.*, 1944, **6**, 317–331.
29. Kaback, G. R. *Vocational personalities: an application of the Rorschach group method.* New York: Bureau of Publications, Teachers Coll., Columbia Univ., 1946.
30. Kelley, D. M., Margulies, H., & Barrera, S. E. The stability of the Rorschach method as demonstrated in electric convulsive therapy cases. *Rorschach Res. Exch.*, **5**, 1941, 35–43.
31. Krugman, J. I. A clinical validation of the Rorschach with problem children. *Rorschach Res. Exch.*, 1942, **6**, 61–70.
32. Krugman, M. Psychosomatic study of fifty stuttering children. *Amer. J. Orthopsychiat.*, 1946, **16**, 127–133.
33. Kurtz, A. K. A research test of the Rorschach test. *Personnel Psychol.*, 1948, **1**, 41–51.
34. Leverett, H. M. Table of mean deviates for various portions of the unit normal distribution. *Psychometrika*, 1947, **12**, 141–152.
35. Lindquist, E. F. *A first course in statistics.* (Revised Ed.) Boston: Houghton Mifflin, 1942.
36. McCandless, B. R. The Rorschach as a predictor of academic success. *J. appl. Psychol.*, 1949, **33**, 43–50.
37. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 1947, **12**, 153–157.
38. Margulies, H. Rorschach responses of successful and unsuccessful students. *Arch. Psychol., N. Y.*, No. 271. New York, 1942.
39. Meltzer, H. Personality differences between stuttering and non-stuttering children. *J. Psychol.*, 1944, **17**, 39–59.
40. Montalto, F. D. An application of the Group Rorschach technique to the problem of achievement in college. *J. clin. Psychol.*, 1946, **2**, 254–260.
41. Munroe, Ruth L. Objective methods and the Rorschach blots. *Rorschach Res. Exch.*, 1945, **9**, 59–73.
42. Munroe, Ruth L. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. *Appl. Psychol. Monogr.*, 1945, No. 7.
43. Munroe, Ruth L. Rorschach findings on college students showing different constellations of subscores on the A.C.E. *J. consult. Psychol.*, 1946, **10**, 301–316.
44. Peatman, J. G. *Descriptive and sampling statistics.* New York: Harper, 1947.
45. Penrose, L. S. Some notes on discrimination. *Ann. Eugenics*, 1947, **13**, 228–237.
46. Piotrowski, Z., Candee, B., Balinsky, B., Holtzberg, S., & Von Arnold. B. Rorschach signs in the

selection of outstanding young male mechanical workers. *J. Psychol.*, 1944, **18**, 131–150.

47. Rapaport, D. *Diagnostic psychological testing, Vol. II.* Chicago: Year Book Publishers, 1946.
48. Richardson, L. H. The personality of stutterers. *Psychol. Monogr.*, 1944, **56**, No. 7.
49. Rickers-Ovsiankina, M. The Rorschach test as applied to normal and schizophrenic subjects. *Brit. J. med. Psychol.*, 1938, **17**, 227–257.
50. Ross, W. D. The contribution of the Rorschach method to clinical diagnosis. *J. ment. Sci.*, 1941, **87**, 331–348.
51. Ross, W. D., Ferguson, G. A., & Chalke, F. C. R. The Group Rorschach Test in officer selection. *Bull. Canad. Psychol. Assn.*, 1945, 84–86.
52. Ross, W. D., & Ross, S. Some Rorschach ratings of clinical value. *Rorschach Res. Exch.*, 1944, **8**, 1–9.
53. Sarbin, T. R., & Madow, L. W. Predicting the depth of hypnosis by means of the Rorschach test. *Amer. J. Orthopsychiat.*, 1942, **12**, 268–271.
54. Schmidt, H. O. Test profiles as a diagnostic aid: the Rorschach. *J. clin. Psychol.*, 1945, **1**, 222–227.
55. Siegel, M. G. The diagnostic and prognostic validity of the Rorschach test in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1948, **18**, 119–133.
56. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Iowa State College Press, 1940.
57. Swift, J. W. Reliability of Rorschach scoring categories with preschool children. *Child Develpm.*, 1944, **15**, 207–216.
58. Swift, J. W. Rorschach responses of 82 pre-school children. *Rorschach Res. Exch.*, 1945, **9**, 74–84.
59. Swineford, F. A table for estimating the significance of the difference between correlated percentages. *Psychometrika*, 1948, **13**, 23–25.
60. Thompson, G. M. College grades and the Group Rorschach. *J. appl. Psychol.*, 1948, **32**, 398–407.
61. Thornton, G. R., & Guilford, J. P. The reliability and meaning of Erlebnistypus scores on the Rorschach test. *J. abnorm. soc. Psychol.*, 1936, **31**, 324–330.
62. Troup, E. A comparative study by means of the Rorschach method of personality development in twenty pairs of identical twins. *Genet. psychol. Monogr.*, 1938, **20**, 461–556.
63. Tulchin, S., & Levy, D. Rorschach test differences in a group of Spanish and English refugee children. *Amer. J. Orthopsychiat.*, 1945, **15**, 361–368.
64. Vaughn, J., & Krug, O. The analytic character of the Rorschach inkblot test. *Amer. J. Orthopsychiat.*, 1938, **8**, 220–229.
65. Vernon, P. E. The Rorschach inkblot test. *Brit. J. med. Psychol.*, 1933, **13**, 179–200.
66. Vernon, P. E. The matching method applied to investigations of personality. *Psychol. Bull.*, 1936, **33**, 149–177.
67. Walker, Helen M. *Elementary statistical methods.* New York: Holt, 1943.
68. Werner, H. Perceptual behavior of brain-injured, mentally defective children. *Genet. Psychol. Monogr.*, 1945, **31**, 51–110.
69. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.
70. Yates, F. The analysis of contingency tables with groupings based on quantitative characters. *Biometrika*, 1948, **35**, 176–181.

HSU, FRANCIS, L. K. *Under the ancestors' shadow: Chinese culture and personality.* New York: Columbia University Press, 1948. Pp. xiv+317. $3.75

Under the Ancestors' Shadow is an anthropological (or sociological) study of "West Town," a small semi-rural community in southwest China. The author expresses his conviction that "the essential social structure" of West Town is typical of China as a whole, although he admits that this judgment can only be confirmed through further study.

The book is appropriately titled: Hsu's account of West Town culture is in some respects far from adequate (in the judgment of a social psychologist), but he does a highly effective job of giving substance and credibility to his central theme as expressed in the title. Taking up one after another aspect of the culture, he elaborates the picture of a people who live "under the ancestors' shadow," and traces the culture patterns which flow from this way of life: the cult of ancestors, family unity, submission to authority, father-son identification, filial piety, and so on.

The economic bases of life in West Town are inexpertly treated. Processes of cultural change, though mentioned a number of times in passing, are dealt with in perfunctory fashion. But for readers of psychological background the chief inadequacies of the book stem from the fact that the author is first and foremost a careful recorder of what Linton has called the *overt* aspects of the culture, and is not particularly skillful in getting at the psychological variables which underlie these overt aspects. There are two shortcomings involved: first, he does not seem skilled at gathering the kinds of data which would permit inferences concerning the psychological variables; and second, when he has such data available, he does not always interpret them with a sure hand. The fact that the book is sub-titled "Chinese Culture and Personality" leads the reader to expect a study in the manner of Linton, Kluckhohn, or Dubois, rich in materials of interest to the psychologist. It falls short of the mark. In dealing with personality, Hsu writes like a man who has been handed his conceptual tools after he had completed his field study.

There is a good deal of internal evidence to indicate that Hsu is a conscientious observer and recorder; but on the whole the study cannot be commended on methodological grounds. Hsu fails to employ the various simple forms of controlled or systematic observation which would have been readily available to him. His treatment of personality structure would have gained immensely from the inclusion of a few autobiographies. He offers little or no information as to his informants. (The reviewer can see no reason why such ethnological field reports as this should not include as a matter of course an "essay on informants," no matter how unpretentious: What sorts of data were obtained from direct observation and what from informants? What kinds of persons supplied what kinds of information? What kinds of persons proved the best and most reliable sources of information? What sorts of resistance and detectable falsifying appeared in informant reports?)

In spite of these shortcomings, *Under the Ancestors' Shadow* is a useful addition to the small but growing shelf of field reports on Chinese culture.

John W. Gardner.

Carnegie Corporation of New York.

BOOKS AND MATERIALS RECEIVED

Anderson, John E. *The psychology of development and personal adjustment.* New York: Henry Holt, 1949. Pp. xvi+720. $3.25.

Cavan, Ruth, *et al. Personal adjustment in old age.* Chicago: Science Research Associates, 1949. Pp. xiii+204. $2.95.

Dennis, Wayne. *Readings in general psychology.* New York: Prentice-Hall, 1949. Pp. xi+525. $3.75.

Donahue, Wilma T., *et al. The measurement of student adjustment and achievement.* Ann Arbor: Univ. of Michigan Press, 1949. Pp. 256. $3.75 cloth. $3.00 paper.

Fetterman, Joseph L. *Practical lessons in psychiatry.* Springfield, Ill.: Charles C Thomas, 1949. Pp. ix+342. $5.75.

Hoch, Paul, and Zubin, Joseph. *Psychological development in health and disease.* New York: Grune & Stratton, 1949. Pp. viii+283. $4.50.

Hovland, Carl I., Lumsdaine, Arthur A., and Sheffield, Fred D. *The American Soldier, Experiments on mass communication.* Vol. III. Princeton: Princeton Univ. Press, 1949. Pp. x+345. $5.00.

Humphrey, George, *Directed thinking.* New York: Dodd, Mead, 1949. Pp. 229. $3.50.

Jenkins, G. G., Shacter, H., and Bauer, W. W. *These are our children.* Chicago: Scott, Foresman, 1949. Pp. 192 (200 photographs). $3.50.

Johnson, Palmer O. *Statistical methods in research.* New York: Prentice-Hall, 1949. Pp. xvi+377.

Jones, Harold E. *Motor performance and growth.* Berkeley: Univ. of California Press, 1949. Pp. x+181. $3.00.

Kanner, Leo. *A miniature textbook of feeblemindedness.* Child Care Monographs No. 1. New York: Child Care Publications, 1949. Pp. 33. $1.75.

LaPiere, R. T., and Farnsworth, P. R. *Social psychology.* (3rd Ed.) New York: McGraw-Hill, 1949. Pp. xiii+626. $4.50.

LINDESMITH, A. R., AND STRAUSS, A. L. *Social psychology.* New York: Dryden Press, 1949. Pp. xvi+549. $4.50.

LIPPITT, RONALD. *Training in community relations.* New York: Harper, 1949. Pp. xiv+286. $3.50.

McKINNEY, FRED. *The psychology of personal adjustment.* (2nd Ed.) New York: John Wiley, 1949. Pp. xi+752. $6.00.

MATHEWSON, ROBERT H. *Guidance policy and practice.* New York: Harper, 1949. Pp. xiv+294. $3.00.

MOON, G. R. *How to become a doctor.* Philadelphia: Blakiston, 1949. Pp. ix+131. $2.00.

MURSELL, JAMES L. *Psychological testing.* (2nd Ed.) New York: Longmans, Green, 1949. Pp. xvi+473. $4.00.

PASTORE, NICHOLAS. *The nature-nurture controversy.* New York: Columbia Univ. Press, 1949. Pp. xvi+213. $3.25.

SANDERS, D. C. *Student selection and academic success.* Sidney, Australia: T. H. Tennant, 1948. Pp. 158.

STENDLER, CELIA B. *Children of Brasstown.* Urbana: Univ. of Illinois Press, 1941. Pp. 103.

SUPER, DONALD E. *Appraising vocational fitness.* New York: Harper, 1949. Pp. xxi+727.

UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949. Pp. vii+638. $4.50.

WALKER, K., AND STRAUSS, E. B. *Sexual disorders in the male.* (3rd Ed.) Baltimore: Williams & Wilkins, 1948. Pp. xiii+269. $3.50.

WARNER, L. W., MEEKER, M., AND EELLS, K. *Social class in America.* Chicago: Social Science Research Council, 1949. Pp. xiii+274. $4.25.

Vol. 46, No. 6 November, 1949

Psychological Bulletin

THE USE AND MISUSE OF THE CHI-SQUARE TEST

DON LEWIS
State University of Iowa

AND

C. J. BURKE
Indiana University

It has become increasingly apparent over a period of several years that psychologists, taken in the aggregate, employ the chi-square test incorrectly. The number of applications of the test does not seem to be increasing, but the number of misapplications does. This paper has been prepared in hopes of counteracting the trend. Its specific aims are to show the weaknesses in various applications that have been made and to set forth clearly the circumstances under which χ^2 can be legitimately applied in testing different hypotheses.

To confirm a general impression that the number of misuses of χ^2 has become surprisingly large, a careful survey was made of all papers published in the *Journal of Experimental Psychology*[1] during the three years 1944, 1945, and 1946. Fourteen papers were found which contained one or more applications of the chi-square test. The applications in only three of these papers (**1**, **15**, **18**) were judged to be acceptable. In one other paper (**21**), the several applications could be called "correct in principle" but they involved extremely small theoretical frequencies. In nine of the fourteen papers (**2**, **4**, **10**, **16**, **17**, **20**, **28**, **29**, **30**) the applications were clearly unwarranted. In the remaining case,[2] it was not possible to determine what had been done; and the author, when questioned twice by letter, did not choose to reply.

The principal sources of error (or inaccuracy) in the fourteen papers just referred to, as well as in papers published in other journals, are as follows:

---

[1] The choice of this particular journal resulted from a belief that the psychologists who publish in it are probably better versed, on the average, in statistical methodologies than are those publishing in other journals. No criticism of the journal nor of individual authors nor of experimental findings is intended. The sole purpose is to illustrate correct and incorrect applications of the chi-square test.

[2] This paper, by Pronko (27), fails to provide the reader with any basis whatever for forming an independent judgment relative to the correctness or incorrectness of the two applications of χ^2 which were made. In this respect it is a good example of the current trend in papers published in psychological journals to reduce explanations of methods of analyzing data to a point where they are quite unintelligible.

1. Lack of independence among the single events or measures[3]
2. Small theoretical frequencies
3. Neglect of frequencies of non-occurrence
4. Failure to equalize ΣF_o (the sum of the observed frequencies) and ΣF_t (the sum of the theoretical frequencies)
5. Indeterminate theoretical frequencies
6. Incorrect or questionable categorizing
7. Use of non-frequency data
8. Incorrect determination of the number of degrees of freedom
9. Incorrect computations (including a failure to weight by N when proportions instead of frequencies are used in the calculations)

These errors will be explained in detail and illustrated with examples taken for the most part from books and published papers.

It is not surprising that errors of the types listed are frequently made; several of the standard texts to which psychologists turn for statistical guidance contain faulty illustrations. For example, Peters and Van Voorhis (**26**) make four applications of χ^2, only one of which is without flaws; and in one of the applications made by Guilford (**13**, p. 91), there is a failure to equalize ΣF_o and ΣF_t and to calculate the number of degrees of freedom correctly.

A single application made by Peters and Van Voorhis contains the first four errors in the above list. Table I is based on their Table Table XXXV (**26**, p. **411**). Twelve dice were thrown fourteen times, and a record was kept of the number of aces appearing at each throw. The observed frequencies F_o are entered in the second column of the table. A value of χ^2, given in the last column, was calculated for each of the fourteen throws. A composite value of χ^2 was obtained by summing the separate values. The first of the four errors in this application is that the observed frequencies lack independence. They lack independence because the same twelve dice were thrown each time. This means that, when the frequencies are grouped, it is impossible to take into account the effects of individual differences in the dice and possible compensating effects from one die to another. As a consequence, no statements can be made about the behavior of an individual die, nor is it possible to generalize the findings to any population of dice from which the twelve can be considered a sample. Therefore, only hypotheses which relate specifically to the twelve dice as a group can be tested. More will be said later about this kind of error.

---

[3] The term independence, as here used, has reference to individual or single events. In contrast, the hypothesis of independence that is tested by means of χ^2 specifies a lack of relationship (that is, an absence of interaction) between the variates represented in a contingency table. The events that occur to yield the frequencies of a contingency table must be mutually independent even though the variates are related.

The second flaw in the application comes from using theoretical frequencies of 2. These values are too low to yield a quantity whose distribution approximates the χ^2 distribution. The third mistake is the failure to equalize ΣF_o and ΣF_t, which are shown in the table as 30 and 28 respectively. This mistake is related to a fourth one—a failure to take account of the frequencies of non-occurrence of aces. Any of the four errors is sufficient to invalidate this use of the χ^2 test.[4]

TABLE I

APPLICATION OF THE CHI-SQUARE TEST BY PETERS AND VAN VOORHIS (**26**)

Throw	F_o	F_t	$(F_o-F_t)^2$	$\frac{(F_o-F_t)^2}{F_t}$
1	1	2	1	0.5
2	3	2	1	0.5
3	2	2	0	0.0
4	3	2	1	0.5
5	1	2	1	0.5
6	4	2	4	2.0
7	2	2	0	0.0
8	4	2	4	2.0
9	1	2	1	0.5
10	0	2	4	2.0
11	3	2	1	0.5
12	2	2	0	0.0
13	3	2	1	0.5
14	1	2	1	0.5
Σ	30	28		10

---

[4] If the observed and theoretical frequencies of non-occurrence of aces had been used in the calculations, the composite value of χ^2 would have been 12 instead of 10. This difference happens not to be large. But in another illustration used by Peters and Van Voorhis (**26**, Table XXXVI, p. 414), the difference is large. The value of 14.52 is given in the text. When χ^2 is correctly computed by taking account of the frequencies of non-occurrence, the resulting value is 29.72. The number of degrees of freedom remains the same, but the calculated value of χ^2 is more than doubled.

FUNDAMENTAL THEORY

The two most basic requirements in any application of the chi-square test are (a) independence among the separate measures and (b) theoretical frequencies of reasonable size. These requirements can be shown in an elementary way by examining a two-category distribution of measures. But first, an unequivocal definition of χ^2 is needed.

If z is a normal deviate in standard form defined in relation to population parameters m and σ, then

$$z = \frac{X - m}{\sigma}$$

and

$$\chi^2 = z^2 = \frac{(X - m)^2}{\sigma^2}, \text{ with } df = 1. \qquad [1]^5$$

Chi-square with 1 degree of freedom is thus defined as the square of a deviation from the population mean divided by the population variance.

If there are r *independent* measures of the variate X, there will be r independent values of z, and the resulting formula for χ^2 is

$$\chi^2 = \sum_{i=1}^{r} z_i^2 = \sum \frac{(X_i - m)^2}{\sigma^2}, \text{ with } df = r. \qquad [2]$$

Values of χ^2 may range from 0 to ∞, and they have frequency distributions which depend upon the value of r. The distribution function of χ^2, in general form, will be given later. It will then be made clear that the chi-square tests of independence and goodness of fit can be applied unequivocally only to frequency data (or to proportions derived from frequency data). For the present, the plausibility of the two basic requirements stated above will be shown through an examination of the two-category case.

The two-category case. Consider a population of N *independent* events (things; measures), each of which may fall either into category A or into category B. It is assumed that these categories are clearly defined before samples are drawn and that the category in which a given event falls can be unequivocally determined. A sample is drawn from the population and the sample data are to be employed in determining whether or not a certain hypothesis regarding the proportion of cases in each category is tenable. If p is the expected (theoretical) proportion

---

[5] Equation [1] verifies the statement that the square root of χ^2 with $df=1$ is distributed as z (or Student's t) with $df=\infty$.

for category A and q the expected proportion for category B, it follows that

$$p + q = 1.$$

The probability $P(n)$ that n of the N events will fall into category A is given by the binomial distribution function

$$P(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n}, \qquad [3]$$

which is the general expression for obtaining the successive terms arising from the expansion of the binomial $(p+q)^N$. The limiting form of equation [3], as N becomes indefinitely large, is a normal distribution function having a mean of Np and a variance of Npq. In symbols,

$$\lim_{N\to\infty} P(n) = \frac{1}{\sqrt{2\pi}\sqrt{Npq}} e^{-(n-Np)^2/2Npq}. \qquad [4]^6$$

Np is the *population* mean of category A. It is the expected value of n, that is, the theoretical frequency to be associated with category A. Nq is the corresponding theoretical frequency to be associated with category B. Npq is the *population* variance. As stated in equation [1], the square of a deviation from the population mean divided by the population variance is distributed as χ^2 with 1 *df*. Thus, from equation [4], it is seen that the quantity

$$\chi^2 = \frac{(n - Np)^2}{Npq} \qquad [5]$$

would be distributed exactly as χ^2 with 1 *df*, provided that N is indefinitely large. In the two-category case, equation [5] may be employed to calculate an approximate value of χ^2. It should be noted that both p and q appear in the denominator of the right-hand term. No restriction is placed during the calculation; the equation gives an approximate solution for any value of N.

The formula that is commonly used in the two-category case to obtain a value of χ^2 is

$$\chi'^2 = \frac{(n - Np)^2}{Np} + \frac{(N - n - Nq)^2}{Nq}, \qquad [6]$$

where Np and Nq are theoretical frequencies and n and $(N-n)$ are the

---

[6] If the investigator is concerned with the probability $P(N-n)$ that $(N-n)$ events will fall into category B, the limiting form of equation [3] would be written

$$\lim_{N\to\infty} P(N-n) = \frac{1}{\sqrt{2\pi}\sqrt{Npq}} e^{-(N-n-Nq)^2/2Npq}. \qquad 4a]$$

corresponding observed frequencies. The prime symbol is placed on χ^2 in [6] to distinguish between χ^2 as *defined* by equation [5] and χ^2 as ordinarily *calculated* with formula [6]. When [6] is used, the number of *df* is 1 less than the number of categories because one restriction ($\Sigma F_t = N$) is imposed on the theoretical frequencies. Therefore, the number of *df* for [6] is 1, just as it is for equation [5].

It can readily be shown that χ^2 and χ'^2 are identical quantities. It is for this reason, and this reason alone, that formula [6] may be used in obtaining an estimate of χ^2 in the two-category case.[7]

The foregoing discussion reveals the two limitations that hold in any application of the chi-square test. The first limitation is that χ^2 is correctly used only if the N events or measures are independent. Equations [3] and [4] are valid statements only when independence exists. The second basic limitation relates to the size of the theoretical frequencies. If Np (or Nq) remains small as N becomes large, the limiting form of the binomial distribution function is not a normal distribution, as was assumed in writing equation [4]. If, for any reason, either Np or Nq is small, the limiting form of equation [3] is the Poisson distribution function. Under such circumstances, the quantity on the right of equation [5] would not be the square of a normal deviate divided by the population variance and, consequently, would not be distributed as χ^2 with 1 *df*.

It should be emphasized that the categories are assumed to be designated in the population before the individual sample is drawn. It should also be emphasized that the equating of the sums of observed and theoretical frequencies *and* the use of the frequency of non-occurrence (in this case, the frequency with which measures fall in category B) are necessary to establish the identity between the quantities defined in equation [5] and equation [6].

---

[7] If observed and theoretical proportions are used in calculating values of χ^2, equations [5] and [6] become

$$\chi^2 = N\frac{(p_o - p_t)^2}{p_t q_t} \qquad [5a]$$

or

$$\chi^2 = N\frac{(q_o - q_t)^2}{p_t q_t}; \qquad [5b]$$

and

$$\chi^2 = N\frac{(p_o - p_t)^2}{p_t} + N\frac{(q_o - q_t)^2}{q_t}. \qquad [6a]$$

In these equations, N is the total number of cases while p_o and q_o are the observed proportions and p_t and q_t are the theoretical proportions for categories A and B, respectively. The equations may be derived from [5] and [6], and they reveal that if proportions are used instead of frequencies, the values calculated from the proportions must be multiplied by N.

The more general case. It is common to have frequency data that fall into several categories instead of just two. This fact requires an extension of the ideas discussed in the two-category case to encompass any number of categories. The basic features of this extension will now be presented. Actually, no new ideas enter into the development. The proof is mathematically more complex, but the underlying ideas are the same.

Consider a population of N *independent* events, with k possible outcomes, $v_1, v_2, v_3 \cdots v_k$. Assume that, in the population,

v_1 occurs with a probability of p_1
v_2 occurs with a probability of p_2
v_3 occurs with a probability of p_3
. . . .
v_k occurs with a probability of p_k

The *joint* probability $P(n_i)$ that out of N events exactly n_1 will fall in category v_1, n_2 will fall in category v_2, n_3 in v_3, *and* $\cdots n_k$ in v_k, is given by the multinomial distribution function

$$P(n_i) = \frac{N!}{n_1!n_2!n_3! \cdots n_k!} p_1^{n_1} p_2^{n_2} p_3^{n_3} \cdots p_k^{n_k}, \qquad [7]$$

where

$$\sum_{i=1}^{k} n_i = N.$$

This is the fundamental expression from which the distribution function of χ^2 is derived. It confirms the statement that the measures (frequencies) in the various cells (categories, classes, etc.) of a multidimensional table must be mutually independent to enable a legitimate application of χ^2 in testing any hypothesis concerning the table. And because equation [7] is written in terms of the frequencies $n_1, n_2, n_3 \cdots n_k$, the chi-square tests of independence and goodness of fit, based as they are on a distribution function derived from [7], may be used unequivocally only in relation to frequency data.[8]

The distribution function of χ^2, here symbolized by $g_r(\chi^2)$, may be written

$$g_r(\chi^2) = C_r(\chi^2)^{(r+2)/2} e^{-(\chi^2/2)}, \qquad [8]$$

where C_r, a coefficient which changes with r, is given by

---

[8] This statement limiting the use of χ^2 to frequency data is not meant to exclude certain special applications such as finding confidence limits for a population variance from a known sample variance or testing several sample variances for homogeneity. These special applications are mentioned again toward the end of the paper.

$$C_r = \frac{1}{2^{(r/2)}\,\Gamma(r/2)}\,. \qquad [9]^{9}$$

It should be emphasized that equation [8] is an exact distribution function for the quantity defined in equations [2] and [5] but only an approximation of the distribution of the quantity defined by equation [7]. Its used in relation to equation [7] requires three separate approximations, each of which assumes a theoretical frequency of reasonable size. The three approximations are:

1. Replacing each of the factorials in equation [7] by its Stirling approximation.
2. Taking a step similar to the one whereby $(1+[X/n])^n$ is replaced by e^X when n is large.
3. Substituting a continuous integral for a summation of discrete quantities.

All of these approximations are quite acceptable and lead to inconsequential errors so long as Np is reasonably large. This reaffirms the fundamental requirement that *the theoretical frequencies must not be small*, if any calculated value of χ^2 is to be distributed as χ^2.[10]

General computational formula. The formula that is commonly employed in calculating values of χ^2 is

$$\chi^2 = \sum \frac{(F_o - F_t)^2}{F_t}, \qquad [10]$$

---

[9] This expression for C_r contains the gamma function $\Gamma(r/2)$. A gamma function is a function which reduces to a factorial whenever the argument is an integer. In the general case, $(n-1)! = \Gamma(n)$. Equation [9] may be written

$$C_r = \frac{1}{2^{(r/2)}\left(\dfrac{r-2}{2}\right)!}\,. \qquad [9a]$$

Whenever r is an even number, the factorial in [9a] is an integral number, and its value along with the value of C_r, can be determined in a straight forward manner. On the other hand, if r is an odd number, the "factorial" is fractional, and its value must be determined either by referring to a table of the gamma function or by using the equation

$$\left(\frac{r-2}{2}\right)! = \left(\frac{r-2}{2}\right)\left(\frac{r-4}{2}\right)\left(\frac{r-6}{2}\right)\cdots\left(\frac{r-[r-1]}{2}\right)(\sqrt{\pi}).$$

A reader who is interested in plots of the χ^2 distribution function for various values of r is referred to Lewis (**19**).

[10] A derivation of equation [8] from equation [7] is presented in considerable detail by Greenhood (**11**) and can be followed by persons familiar with advanced calculus. Greenhood's development, which is more complete but similar to one given by Fry (**8**), indicates clearly the limitations which hold in applications of the chi-square test to frequency data. It is not necessary, of course, to derive equation [8] from the multinomial distribution function; it may be derived directly from the joint normal distribution function in n variables.

where F_o and F_t, as usual, are observed and theoretical frequencies and the summation extends over all cells (categories) of the table. The equation holds for any number of categories and reduces to the form of [6] when the number of categories is 2. It serves as a constant reminder that the chi-square tests of independence and goodness of fit can be applied unequivocally only to frequency data.

APPLICATIONS: I. THE GOODNESS OF FIT OF DISTRIBUTION FUNCTIONS

One of the commonest applications of the chi-square test is in evaluating the hypothesis that a set of frequency data can be satisfactorily represented by some specified distribution function. It makes little difference what the function is, so long as its fundamental properties are known. The goodness of fit of binomial, Poisson, and normal distribution functions is often tested.

Two-category case. The correct use of χ^2 in connection with a symmetrical binomial distribution function (where $p=q$) can be illustrated with data obtained in a coin-guessing "experiment." A coin was tossed, and 96 students of elementary psychology each guessed whether the coin came up "heads" or "tails." The hypothesis to be tested is that the guess of each student, like the fall of the coin itself, was a purely chance occurrence and that each student was as likely to say heads as to say tails. The results are shown in Table II.

TABLE II

DATA FROM A COIN-GUESSING EXPERIMENT

H_o ($=F_o$ for heads)	68
H_t ($=F_t$ for heads)	48
T_o ($=F_o$ for tails)	28
T_t ($=F_t$ for tails)	48

As seen in the first and third rows, 68 students guessed heads and 28 guessed tails. The theoretical frequencies are 48 and 48. The value of χ^2 is obtained as follows:

$$\chi^2 = \frac{(H_o - H_t)^2}{H_t} + \frac{(T_o - T_t)^2}{T_t} = \frac{(68-48)^2}{48} + \frac{(28-48)^2}{48} = 16.67.^{11}$$

With 1 *df*, this value is significant at better than the 0.1% level of confidence;

---

[11] Two alternative formulas that may be used in the two-category case when $p=q$ are as follows:

$$\chi^2 = (H_o - H_t)^2(2/H_t);$$

$$\chi^2 = \frac{(H_o - T_o)^2}{H_o + T_o}.$$

These formulas are the exact equivalents of the one used above.

so it must be concluded that the guesses of the 96 students were somehow biased in favor of heads.

Additive property of χ^2, illustrated with two-category data. A fundamental property of χ^2 is indicated by the rule which states that the sum of any number of separate and independent values of χ^2 is distributed as χ^2, the number of *df* being the sum of the separate *df*'s. The rule will now be applied in relation to coin-guessing data. Two applications will be made, the first of which is incorrect. *This incorrect application is purposely included* as a means of showing how it differs from a correct application and also of revealing the source of many of the errors made by investigators when employing the chi-square test.

Ninety-six students of elementary psychology were given five successive "trials" in coin guessing. A single coin was tossed five times. After each toss, the 96 students each guessed whether the coin came up "heads" or "tails." Each student wrote his five guesses in order on a sheet of paper. The turn of the coin was never revealed. The results are summarized in Table III. The frequencies for toss 1 are the same as those given in Table II.

TABLE III

DATA FROM A SECOND COIN-GUESSING EXPERIMENT

	Tosses					
	1	2	3	4	5	
H_o	68	49	39	54	54	
H_t	48	48	48	48	48	
T_o	28	47	57	42	42	
T_t	48	48	48	48	48	
χ^2	16.67	.04	3.37	1.50	1.50	$\sum = 23.08$

With the ten pairs of observed and theoretical frequencies in Table III (two pairs for each toss), it is possible to test five separate hypotheses—that the guesses of the 96 students on *any one* of the five tosses were chance occurrences and were as apt to be heads as tails. Five values of χ^2 are given in the bottom row of the table. Each of the values was computed with the formula employed with the data of Table II. The number of degrees of freedom is 1 in each case. For 1 *df*, the value of χ^2 at the 5% level of confidence is 3.841. Four of the five calculated values are less than this and provide no satisfactory basis for rejecting the hypothesis of chance occurrence.

To secure an estimate of χ^2 which makes the probability of rejecting a false hypothesis large, it is desirable, when conditions warrant, to summate separate estimates of χ^2 and obtain a composite estimate. The number of *df* for the composite estimate is always the sum of the separate *df*'s. The sum of the five values in Table III is 23.08. The number of *df* is 5. The hypothesis under test is that the guesses of students *on five successive tosses* of a coin are purely chance occurrences, with the probability of a guess of heads (by any student on any toss) equal to the probability of a guess of tails. The composite value of

23.08 is significant at better than the 0.1% level of confidence, and would warrant a rejection of the hypothesis if the application of the test were correct.

But it is not correct to summate the five values of χ^2 in Table III. The reason is that the responses of the 96 students from one toss to the next cannot be assumed to have been independent. In other words, the tabulated values of H_o and T_o from toss to toss are interdependent. Consequently, there is no way of obtaining an unbiased estimate of the theoretical frequency for any toss beyond the first (unless previous guesses are completely ignored). It is unreasonable to assume that a student's knowledge of his guess on one toss did not influence his guess on succeeding tosses. (It will be shown from the data that such an assumption is unsound.) The χ^2 test should never be based on an assumption that is already known to be false. On a single toss, the guess of each student was independent of the guess of the other students; but between tosses, the five guesses of each student were undoubtedly interrelated.

A lack of independence between separate events (measures) is the commonest flaw in the applications of χ^2 that are made by psychologists. Six of the fourteen papers (**2**, **10**, **17**, **28**, **29**, **30**) referred to in the opening paragraphs contain applications having this shortcoming.

There are two ways in which the responses of subjects may be interdependent. They may be related from trial to trial, as they were in the coin-guessing illustration, or they may be internally linked within a single trial. Whenever individual subjects each make more than one response per trial, linkages among the measures within the trial must result unless there are no individual differences. Many investigators ignore this restriction and apply the χ^2 test even though the same subjects are used from trial to trial and make several responses on each trial.

The correct use of the additive principle with two-category data. The conditions under which separate values of χ^2 may be legitimately summated can be illustrated with two-category data. Five non-overlapping groups of subjects made a single guess on each of five successive tosses of a coin. The number of subjects per group was 86. Each subject wrote down his guesses in order on a sheet of paper, the sequence of his guesses remaining on the sheet before him. The results for the five groups are summarized in Tables IV (A to E, inclusive). Twenty-five separate values of χ^2 appear in the bottom rows of the five sections of the table. The question to be answered is: *Which of these 25 values may be legitimately summated and which may not be?* The five values in any one of the sections cannot be meaningfully added for reasons already given in the discussion of the data in Table III. However, the five values in the five sections for *any single toss* may be summated to yield a single composite value having 5 *df*. For example, a composite value may be obtained for the third toss and may be used to test the hypothesis that, on the third toss of the coin, the guesses of the members of the five groups were random occurrences with the probability of a heads guess equalling that of a tails guess. Neither the presence nor the absence of individual biases would nullify a meaningful test of this hypothesis. If the hypothesis could not be retained, it would be correct to conclude that the guesses were not chance occurrences and were perhaps influenced by what had gone before. The difference between this case and the one illustrated in Table III is clear-cut. The data in Table III are for five successive guesses by one group, and the successive theoretical probabilities cannot be established

TABLE IV

DATA FROM A THIRD COIN-GUESSING EXPERIMENT

	Tosses				
	1	2	3	4	5
			A. Group 1		
H_o	63	56	36	45	41
H_t	43	43	43	43	43
T_o	23	30	50	41	45
T_t	43	43	43	43	43
χ^2	18.60	7.86	2.28	0.19	0.19
			B. Group 2		
H_o	63	56	42	48	45
H_t	43	43	43	43	43
T_o	23	30	44	38	41
T_t	43	43	43	43	43
χ^2	18.60	7.86	0.05	1.16	0.19
			C. Group 3		
H_o	65	55	40	48	46
H_t	43	43	43	43	43
T_o	21	31	46	38	40
T_t	43	43	43	43	43
χ^2	22.51	6.70	0.42	1.16	0.42
			D. Group 4		
H_o	68	54	38	52	41
H_t	43	43	43	43	43
T_o	18	32	48	34	45
T_t	43	43	43	43	43
χ^2	29.07	5.62	1.16	3.77	0.19
			E. Group 5		
H_o	72	57	30	62	38
H_t	43	43	43	43	43
T_o	14	29	56	24	48
T_t	43	43	43	43	43
χ^2	39.12	9.12	7.86	16.79	1.16

without making specific assumptions concerning prior events. In contrast, when the five χ^2 values for a single toss are taken from the five sections of Table IV and summated, the only assumption that is made is that the guesses on that particular toss were in accordance with a theoretical probability for heads of .50.

Composite values of χ^2 for the five tosses are given in Part A, Table V.

TABLE V

VALUES OF CHI-SQUARE BASED UPON DATA FROM TABLE IV

Part A. Composite Values of χ^2

Toss	χ^2	*df*
1	127.90	5
2	37.16	5
3	11.77	5
4	23.07	5
5	2.15	5

Part B. Values of χ^2 *for Combined Frequencies*

Toss	χ^2	*df*
1	125.17	1
2	36.92	1
3	7.82	1
4	14.88	1
5	.149	1

The number of *df* in each case is 5. Except for toss 5, the composite values are all significant at far better than the 1% level of confidence. The value for toss 5 falls near the 80% level and lends support for the belief that on a fifth successive toss of a coin, the probability of a heads guess is .50. The values for tosses 1 and 2 show that, on these tosses, the guesses were strongly biased toward heads. Conclusions regarding tosses 3 and 4 should be made in the light of the large contributions made to the composite values for these tosses by the guesses of group 5 alone. As seen in Table IV-E, the χ^2 values for tosses 3 and 4 are 7.86 and 16.79, respectively. With 1 *df* in each case, both values are significant at better than the 1% level. The corresponding values in the other four parts of the table all fall below the 5% level. However, the deviations on tosses 3 and 4 are in the same direction for all five groups, and this fact indicates a definite departure from chance expectations.

It is legitimate to combine for each toss separately the empirical and theoretical frequencies listed in Table IV-A-B-C-D-E, and use the resulting sums to compute values of χ^2, each with 1 *df*. Values of χ^2 obtained in this way are given in Part B of Table V. As in the case of the composite values, all of the values except the one for toss 5 are significant at better than the 1% level of confidence. Nevertheless, the procedure of combining frequencies is not recommended except where the theoretical frequencies for each of several duplicated experiments are too small to yield satisfactory individual estimates of

χ^2. Other things equal, the greater the number of degrees of freedom is, the more stable is a value of χ^2 and the greater is the probability of rejecting a false hypothesis.

Five values of χ^2 may be selected from Table IV-A-B-C-D-E and used compositely to test the hypothesis that the guesses of persons on *five successive tosses* of a coin are chance occurrences, with the probability of a heads guess equalling that of a tails guess. The hypothesis is inclusive enough to cover the entire population from which the groups of subjects were randomly selected. To provide for independence between tosses, it is necessary to choose the five values of χ^2 so that there is one value for each toss and so that no two values are based on the guesses of a single group. To this end, numbers from 1 to 5 were assigned to the five sections of the table. The χ^2 value for toss 1 was taken from the section whose number first appeared in a table of random numbers; the χ^2 value for toss 2 from the section whose number next appeared, and so on. The values thus chosen are shown in Table VI. The composite value is 33.66, with 5 *df*. It is highly significant and leaves no grounds for believing that the hypothesis is true.

TABLE VI

COMPOSITE VALUE OF CHI-SQUARE, BASED ON DATA IN TABLE IV

	Tosses				
	1	2	3	4	5
H_o	63	54	30	48	46
H_t	43	43	43	43	43
T_o	23	32	56	38	40
T_t	43	43	43	43	43
χ^2	18.60	5.62	7.86	1.16	0.42
Composite $\chi^2 = 33.66$					

The discussion in the preceding pages on the necessity for independence between measures can perhaps be further clarified through a consideration of coin-tossing. Three different situations will be described to reveal unmistakable differences in hypotheses to be tested and methods of handling data. Suppose, first, that a single penny is selected at random from a large collection of pennies. This penny is tossed successively, say 100 times, and a record is kept of the way it turns. The probability of a head (or a tail) is .50. The χ^2 test may be applied to determine whether or not the empirical results conform to this theoretical probability. If they do conform, it may justifiably be concluded that the penny is "unbiased." The test is an unequivocal one. The extent to which the investigator should generalize to the collection of pennies from which the single penny was chosen is a matter for personal judgment. The χ^2 test, as made, reveals nothing concerning the probability that the selected penny either represents or misrepresents the collection of pennies.

Suppose next that *two* pennies are randomly selected from a collection of pennies and that each penny is tossed 50 times to give a total of 100 tosses.

The frequencies of occurrence of heads and of tails are combined (pooled) for the two coins. It may be assumed that the probability of a head is .50, but the χ^2 test cannot be *meaningfully* applied to test this theoretical probability.[12] The reason is that each penny makes its own unique contribution to the results. If one of them is biased while the other is unbiased, the obtained value of χ^2 could easily be significant and lead to a rejection of the hypothetical probability, even though it is correct for one of the pennies. Furthermore, one penny could be strongly biased for heads and the other equally strongly biased for tails, and the obtained value of χ^2 would turn out to be insignificant. The possible presence of individual idiosyncrasies precludes an unequivocal application of χ^2. The same thing would be true if five pennies were randomly selected, each one tossed, say 20 times, and the results pooled; or if 10 pennies were selected and each tossed 10 times, or 20 selected and each tossed five times.

Suppose, finally, that 100 pennies are randomly selected from a collection of pennies, that each penny is tossed a single time, and that the number of heads is recorded. It may be assumed that the probability of a turn of heads, in the population from which the pennies are selected, is .50. The fall of each coin is clearly independent of the fall of every other coin. The χ^2 test may be legitimately applied to determine whether or not the observed frequency of heads conforms to the hypothetical frequency. The results of the test can be generalized to the entire collection of pennies. This would hold even though less than 100 pennies were selected, so long as a sufficient number was chosen to provide theoretical frequencies of the occurrence of heads and the occurrence of tails of sufficient magnitude to warrant an application of the χ^2 test. No statements can be made, of course, regarding the tendencies of any individual penny.

The crucial point is that frequencies obtained from individuals, whether pennies or subjects in psychological experiments, should not be pooled if the χ^2 test is to be used, except when it can be shown that there is an absence of biases or idiosyncrasies among them[13] or when "interaction" effects are specifically under scrutiny. Results on individuals may be combined, but the combining should be done *after* the χ^2 test has been applied to the data on individuals separately. For example, if two pennies are each tossed 50 times, the χ^2 test may be applied to the results for each penny separately, and then the two values of χ^2 may be added to provide a composite value. Similarly, separate values of χ^2 may be obtained from the guesses made by two individuals. The separate values may then be combined to furnish a single composite value. As in the well-known analysis of variance techniques where each source of variability contributes to the total variability, each source of variability should be allowed to make its contribution to the value of χ^2. Unfortunately, χ^2 procedures provide no way, as analysis of variance techniques do, of introducing *statistical* controls over individual subjects as a source of variability.

---

[12] The pooling of two or more sets of frequencies to obtain a single value of χ^2 is warranted if the aim is to study the "heterogeneity" or "interaction" aspects of the data. In this connection, see Snedecor's discussion (**31**, pp. 191–192) of "pooled" and "total" chi-squares.

[13] If an investigator firmly intends to restrict all generalizations to the group of persons studied—the group considered *in toto*, as a sort of amorphous mass—then the pooling of individual frequencies may be logically defended. In such a situation, the group is analogous to a single individual and must be treated as such.

Therefore, in the use of χ^2, the control over individuals must be introduced as an intrinsic part of the sampling process.

Multi-category case (single dimension). Frequency data sometimes fall into several categories along a single dimension. If the frequencies from category to category are independent and if some hypothesis regarding their distribution can be meaningfully set up, then the chi-square test may be used to evaluate the hypothesis, provided that the theoretical frequencies for the various categories are of reasonable magnitude. The correct application of the test in such a situation will be illustrated with data obtained in die throwing. A single die was thrown 120 times. There was no reason for believing that any throw was influenced by any other. The results are given in Table VII, where the first column lists the six faces of the die and the second column gives the number of times that each face turned up. The theoretical probability, on each throw, that any specified face of the die would turn up was 1/6. Consequently, the theoretical frequency of occurrence of each of the six faces was 20, as shown in the third column of the table. A single restriction, the sum of the observed frequencies, was placed in calculating the theoretical frequencies. Consequently, there are 5 *df*. With 5 *df*, the obtained value of χ^2 falls near the 80% level of confidence, and there is no basis for rejecting the hypothesis that the fall of the die was, on each throw, a strictly chance occurrence.[14]

TABLE VII

DATA OBTAINED IN DIE THROWING

Face	F_o	F_t	$\frac{(F_o - F_t)^2}{F_t}$
1	23	20	.45
2	20	20	.00
3	22	20	.20
4	15	20	1.25
5	18	20	.20
6	22	20	.20
$\sum$	120	120	2.30

---

[14] It would be possible to use the data in Table VII to test six separate hypotheses—that the appearance of *each* of the six faces was a chance occurrence, with a probability of 1/6. For example, to test the hypothesis that the appearance of the ace (one-spot) was a chance occurrence, a value of χ^2 would be computed as follows:

$$\chi^2 = \frac{(23-20)^2}{20} + \frac{(97-100)^2}{100} = .45 + .09 = .54.$$

The number of *df* is 1. Observe that in the calculation, the frequency of non-occurrence of the ace was taken into account. (The probability of occurrence of the ace was 1/6 [$=p$] while the probability of its non-occurrence was 5/6 [$=q$]. This is an example of an asymmetrical binomial.) It is clear that six separate values of χ^2 could be computed

Another illustration of a multi-category frequency distribution (along a single dimension) comes from results on the coin-guessing experiment. A total of 439 subjects each guessed heads or tails on each of five successive tosses of a coin. All but 9 of the subjects were the ones whose guesses were tabulated in Table IV. The χ^2 test will be applied to evaluate the hypothesis that chance factors operated in determining the frequencies of occurrence of the various possible "patterns" of guesses. With five successive tosses and five successive guesses, there were 32 possible patterns, as shown in the first column of Table VIII. If there was no biasing of the guesses (that is, if the guess of every subject on every trial was as likely to be heads as to be tails), then each of the 32 patterns was as probable as any other; and with $N=439$, the theoretical frequency for each pattern was $1/32 \times 439 = 13.7$, rounded to the first decimal place. This is the value shown in the third column of the table, except where parentheses appear. Note that 32×13.7 does not equal 439 exactly, but equals 438.4. One of the requirements in the application of χ^2 is that $\sum F_o = \sum F_t$. Therefore, six of the theoretical frequencies are given as 13.8. These are the theoretical frequencies corresponding to the six largest observed frequencies. This insures that any slight error that may result from equalizing $\sum F_o$ and $\sum F_t$ operates to make the test more conservative.

In the calculation of χ^2 for Table VIII, a value of $(F_o - F_t)^2/F_t$ was secured for each row. These values are given in the fourth column, and their sum ($=681.122$) is the desired estimate. The number of *df* is $(32-1)=31$. The only restriction placed in figuring the theoretical frequencies was $\sum F_o$, and this meant the loss of a single degree of freedom. Even with 31 *df*, the value of χ^2 is so large as to leave no basis whatever for retaining the hypothesis that the guesses were chance occurrences.

Normal distribution function. The chi-square test is often used in evaluating the fit of a normal curve to a set of frequency data. Applications of this type are usually correct except for an occasional failure to equalize $\sum F_o$ and $\sum F_t$, a tendency to use some theoretical frequencies that are too small, and, most importantly, an incorrect specification of the number of degrees of freedom. The correct procedure will now be illustrated.

The distribution of the midterm scores of 486 students in a course in elementary psychology is shown in Table IX. The mid-points of class intervals of ten score units are given in the X-column, frequencies in the F_o-column. The mean M of the distribution of scores is 104.0, while the standard deviation is 16.1. There are two methods that can be used in fitting a normal curve to the data (that is, in calculating the theoretical frequencies that correspond to the observed frequencies). One method involves the estimation of areas under segments of the normal curve through the process of multiplying ordinate values by the class interval.[15] This is an approximation procedure. A more exact method (and the one used here) is to obtain the values for the areas from proportions taken from a table of the probability integral.

---

in the way just indicated. But these values could not then be legitimately summated to yield a composite value of χ^2 with 6 *df*. They could not be combined because they would lack independence; the frequency of non-occurrence in each calculation would include the frequency of occurrence of the other five faces.

[15] For an example of this method, see Guilford (**13**, p. 91).

The column in Table IX labeled X' gives the upper limits of the various score categories. Deviate scores and z scores based on the values of X' are shown in the fourth and fifth columns. Proportions of the total area under the normal curve from $-\infty$ to z are given in the P column. Proportions of the

TABLE VIII

ANALYSIS OF COIN-GUESSING RESPONSES OF 439 SUBJECTS*

Patterns	F_o	F_t	$\frac{(F_o - F_t)^2}{F_t}$
H H H H H	25	(13.8)	9.089
H H H H T	12	13.7	0.211
H H H T H	22	13.7	5.028
H H H T T	18	13.7	1.350
H H T H H	29	(13.8)	16.742
H H T H T	96	(13.8)	489.626
H H T T H	22	(13.8)	4.872
H H T T T	14	13.7	0.007
H T H H H	5	13.7	5.525
H T H H T	15	13.7	0.123
H T H T H	12	13.7	0.211
H T H T T	4	13.7	6.868
H T T H H	33	(13.8)	26.713
H T T H T	17	13.7	0.795
H T T T H	12	13.7	0.211
H T T T T	3	13.7	8.357
T H H H H	3	13.7	8.357
T H H H T	7	13.7	3.277
T H H T H	10	13.7	0.999
T H H T T	14	13.7	0.007
T H T H H	5	13.7	5.525
T H T H T	1	13.7	11.773
T H T T H	6	13.7	4.328
T H T T T	0	13.7	13.700
T T H H H	4	13.7	6.868
T T H H T	6	13.7	4.328
T T H T H	25	(13.8)	9.089
T T H T T	7	13.7	3.277
T T T H H	2	13.7	9.992
T T T H T	5	13.7	5.525
T T T T H	2	13.7	9.992
T T T T T	3	13.7	8.357
$\sum$	439	439.0	$\chi^2 = 681.122$

* All but nine of the subjects are the same as those whose guesses are analyzed in Table IV.

area in the segments corresponding to the various score intervals are shown in column P' and were obtained by taking the differences between the successive values of P. The theoretical frequencies came from multiplying the values of P' by N which is 486 in this case.

TABLE IX

Application of the Chi-Square Test in Evaluating the Fit of a Normal Curve to a Set of Frequency Data

X	F_o	X'	x $(=X'-M)$	z $(=x/\sigma)$	P	P'	F_t $(=P'N)$	$\frac{(F_o-F_t)^2}{F_t}$
				$(+\infty)$	(1.0000)			
144.5	1					.0141	6.9	.926
		139.5	35.5	2.195	.9859			
134.5	22					.0437	21.2	
		129.5	25.5	1.574	.9422			
124.5	56					.1125	54.7	.031
		119.5	15.5	0.953	.8297			
114.5	112					.1960	95.3	2.926
		109.5	5.5	0.342	.6337			
104.5	111					.2440	118.5	.475
		99.5	− 4.5	−0.280	.3897			
94.5	94					.2032	98.8	.233
		89.5	−14.5	−0.891	.1865			
84.5	54					.1211	58.8	.392
		79.5	−24.5	−1.512	.0654			
74.5	27					.0489	23.8	.555
		69.5	−34.5	−2.133	.0165			
64.5	7					.0135	6.6	
		59.5	−44.5	−2.754	.0030			
54.5	2					.0030	1.4	
				$(-\infty)$	(.0000)			
$\sum$	486					1.0000	486.0	5.538

Because the first and the last two values of F_t are less than 10, they were combined with the adjacent values, as were the corresponding values of F_o. The sum of the last column in the table (=5.538) is the value of χ^2. Seven differences between F_o and F_t entered into the calculations. The number of degrees of freedom is $7-3=4$. Three degrees were lost because three restrictions were placed in determining the theoretical frequencies.[16] The restrictions were the computed values of $\sum F_o$, M, and σ. The hypothesis being tested is that the frequency data arose from a normal population. With 4 *df*, the probability of obtaining, by chance, a value of χ^2 greater than 5.538 is around .25; so the hypothesis is tenable.

---

[16] Some statistics texts (6, 25, 26) perpetuate the view, erroneously attributed to Pearson (**24**), that the number of restrictions imposed in fitting a normal curve is *1* or *3*, depending upon the hypothesis that the investigator wishes to test. There is only one hypothesis open to test—that the frequency data arose from a normal population. If

The Poisson distribution function. If the probability of the occurrence of an event is quite small, so that Np remains small even though N is relatively large, the distribution of observed frequencies in samples of size N may be of the Poisson type. The equation for the Poisson distribution may be written

$$P(n) \doteq \frac{m^n}{n!} e^{-m} \qquad [11]^{17}$$

where $m = Np$ and e has its conventional meaning. As in equations [3] and [4], the symbol $P(n)$ represents the probability of n occurrences out of N possible occurrences. The symbol $\doteq$ is used in place of the equal sign to indicate that [11] is an approximation formula. The errors introduced by the approximation are negligible, provided that N is quite large and provided also that N is very much larger than the largest value of n that may reasonably be expected in random sampling.

TABLE X

ANALYSIS OF FREQUENCIES OF OCCURRENCE OF THE CONSONANT "TH" IN SAMPLES OF AMERICAN SPEECH

n	F_o	$P(n)$	F_t	$(F_o-F_t)^2$	$\frac{(F_o-F_t)^2}{F_t}$
0	31	.1868	22.42	73.61	3.283
1	31	.3138	37.66	44.35	1.178
2	30	.2636	31.63	2.66	.084
3	11	.1476	17.71	45.02	2.542
4	11	.0620	7.44	41.22	3.896
5	3	.0209	2.51		
6	3	.0050	.60		
. . .		. . .	.		
>6		.0003	.03		
$\sum$	120	1.0000	120.00		$\chi^2 = 10.983$

The χ^2 test may be applied in relation to the data of Table X. The column labeled F_o gives the number of samples, in a total of 120 short samples of

---

the mean and standard deviation of the fitted function are estimated from the data, three restrictions are imposed and 3 *df* are lost. The same texts give a similar misinterpretation of the number of restrictions imposed when χ^2 is applied in testing independence.

[17] A derivation of this formula is given by Lewis (**19**, pp. 168–169).

American speech, that contained n occurrences of the consonant "th" (as in thin). Each of the 120 samples was 400 sounds in length. As seen from the table, 31 of the samples did not contain any "th" sounds; 31 samples contained one "th" sound each; and so on. As a first step, the data in Table X will be compared with some results obtained by Voelker (33). In a study of over 600,000 sounds occurring in almost 6,000 announcements over the radio, Voelker found the proportion of "th" sounds to be .0065. Each of the 120 samples represented in Table X contained 400 sounds. This made a total of 48,000 sounds. The use of the proportion obtained by Voelker leads to 260 as the predicted, or theoretical, number of "th" sounds among the 48,000. The observed number was 201.[18] The χ^2 test may be applied in evaluating the hypothesis that the sounds in the present over-all sample, were drawn from a general population of American speech sounds which is assumed to be characterized exactly by the value of m obtained by Voelker. The value of χ^2 is computed as follows:

$$\chi^2 = \frac{(201-260)^2}{260} + \frac{(47,799-47,740)^2}{47,740} = 13.388 + 0.073 = 13.461.$$

Note that the observed and theoretical frequencies of non-occurrence of the "th" enter into the calculation. With 1 *df*, the obtained value of χ^2 is significant at better than the 0.1% level of confidence; so there is a firm basis for rejecting the hypothesis.

The observed proportion of .0042 ($=201/48,000$) may be used in testing the hypothesis that the 120 samples were all drawn from the same Poisson distribution. If $p_o=.0042$ and $N=400$, then $m=Np=1.68$, and the equation for the hypothetical distribution function may be written

$$P(n) = \frac{(1.68)^n}{n!}\, e^{-1.68}.$$

Values of $P(n)$ computed with this formula are given in the third column of Table X, and the corresponding values of F_t are given in the fourth column. As shown in the last column of the table, the value of χ^2 was computed in accordance with equation [10]. The last 4 values of F_t and the last 3 values of F_o were combined to avoid the use of theoretical frequencies of less than 10. The computed value of χ^2 is 10.982. Five differences between F_o and F_t were used in the computations. Two restrictions (N and m)[19] were imposed in calculating values of F_t. This leaves 3 *df*. With this number of *df*, the obtained value χ^2 falls at about the 3% level of confidence. The hypothesis may, therefore, be tentatively retained or may be rejected, depending upon the level of confidence that has been prescribed.

APPLICATIONS: II. THE CHI-SQUARE TEST OF INDEPENDENCE

A common application of the χ^2 test enables an examination of the frequencies of a contingency table to determine whether or not the two variables or attributes represented in the table are independent. The number of cells in the table may range from four (as in a 2×2 table) to an indefinitely large value. The χ^2 test is perhaps most commonly applied by psychologists in relation to 2×2 tables. The chief weaknesses in such applications are (a) a strong tendency to use excessively small theoretical frequencies and (b) an occasional failure to categorize adequately. The same two weaknesses are apt to occur when the number of categories in either, or both, of the "dimensions" of the table is greater than two.

---

[18] $\sum nF_o=0(31)+1(31)+2(30)+3(11)+4(11)+5(3)+6(3)=201.$

[19] If the value of p had not been estimated from the empirical data, but had been taken from the Voelker study or from some other completely independent source, the only restriction that would have been placed would have been N, and there would have been 4 *df* instead of 3 *df*.

Illustrations of the Chi-Square Test of Independence

Comparison of coin- and die-coin guessing. To obtain data for illustrating the χ^2 test of independence, 384 students of psychology were each asked to guess heads or tails on five successive tosses of a coin, where the tosses were interspersed among five throws of a die. The die was thrown; a guess was made as to the face that turned up. This guess was written on one edge of a sheet of paper. The edge of the paper was then folded under, to hide the guess. The

TABLE XI

RESULTS FROM COIN- AND DIE-COIN GUESSING EXPERIMENTS

Patterns	F_o			F_t	
	Coin Guessing	*Die-Coin Guessing*	*Totals*	*Coin Guessing*	*Die-Coin Guessing*
H H H H H	25	29	54	28.8	25.2
H H H H T	12	9	21	11.2	9.8
H H H T H	22	9	31	16.5	14.5
H H H T T	18	5	23	12.3	10.7
	—	—			
	77	52			
H H T H H	29	38	67	35.7	31.3
H H T H T	96	59	155	82.7	72.3
H H T T H	22	24	46	24.5	21.5
H H T T T	14	8	22	11.7	10.3
	—	—			
	161	129			
H T H H H	5	7	12	6.4	5.6
H T H H T	15	9	24	12.8	11.2
H T H T H	12	2	14	7.5	6.5
H T H T T	4	9	13	6.9	6.1
	—	—			
	36	27			

TABLE XI—*Continued*

Patterns	F_o			F_t	
	Coin Guessing	*Die-Coin Guessing*	*Totals*	*Coin Guessing*	*Die-Coin Guessing*
H T T H H	33	12	45	24.0	21.0
H T T H T	17	20	37	19.7	17.3
H T T T H	12	3	15	8.0	7.0
H T T T T	3	4	7	3.7	3.3
	—	—			
	65	39			
Sub-totals	(339)	(247)			
T H H H H	3	3	6	3.2	2.8
T H H H T	7	5	12	6.4	5.6
T H H T H	10	9	19	10.1	8.9
T H H T T	14	3	17	9.1	7.9
	—	—			
	34	20			
T H T H H	5	4	9	4.8	4.2
T H T H T	1	8	9	4.8	4.2
T H T T H	6	4	10	5.3	4.7
T H T T T	0	1	1	0.5	0.5
	—	—			
	12	17			
T T H H H	4	13	17	9.1	7.9
T T H H T	6	19	25	13.3	11.7
T T H T H	25	37	62	33.1	28.9
T T H T T	7	13	20	10.7	9.3
	—	—			
	42	82			
T T T H H	2	4	6	3.2	2.8
T T T H T	5	5	10	5.3	4.7
T T T T H	2	1	3	1.6	1.4
T T T T T	3	8	11	5.9	5.1
	—	—			
	12	18			
Sub-totals	(100)	(137)			
Totals	439	384	823	438.8	384.2

coin was then tossed, the guess being written down. Again the paper was folded under, to hide the guess. The die was thrown a second time, the guess made, the paper folded under. The coin was then thrown a second time, the guess made, the paper folded under. Each of the five guesses on the coin was preceded by a throw of the die and a guess on its fall. The paper was folded under after each guess on the die and each guess on the coin. Thus, the guesses on the coin were not only separated by guesses on the die, but the sequence of guesses was hidden from view. The subjects were never informed as to how the coin or the die actually fell.

The frequencies of occurrence of the 32 possible patterns of guesses on the five successive tosses of the coin (with guesses on the die ignored) are given in the third column of Table XI. The frequencies in the second column of this table were copied directly from Table VIII and are based on guesses on five successive tosses of a coin alone.[20] The subjects for the two conditions of guessing were completely different.

TABLE XII

2×2 CONTINGENCY TABLE BASED UPON THE COIN- AND DIE-COIN GUESSING EXPERIMENTS

	First Guess		
	H	*T*	
Coin Guessing	339 (312.6)	100 (126.4)	439
Die-Coin Guessing	247 (273.4)	137 (110.6)	384
	586	237	823

The first 16 patterns listed in Table XI begin with a guess of heads, the last 16 with a guess of tails. A 2×2 contingency table, shown in Table XII, was set up, the division along one "dimension" being between first-guess-heads and first-guess-tails and along the other "dimension" between the two conditions of guessing. The four sub-totals in Table XI constitute the observed frequencies appearing in the four cells of the 2×2 table.

The hypothesis to be tested in this case is that the occurrence of a first guess of heads was independent of the condition under which the guessing was done. If the observed frequencies in Table XII were independent of the conditions of guessing, the probability of a guess of heads was 586/823. The probability of a subject's being in the coin guessing group was 439/823. The joint probability that a subject would be in the coin guessing group and would also guess heads was 439/823 × 586/823. The theoretical frequency for the upper left-hand cell of the table was obtained by multiplying this joint probability by 823. The other three theoretical frequencies were automatically determined by this single calculated frequency and by the restrictions of the border sums. The four values of F_t are shown in parentheses in the table. The value of χ^2 was computed as follows:

---

[20] As will be seen, the 32 patterns in Table XI have been divided into groups of four patterns each. The basis for the division will be discussed later. Short horizontal lines divide the corresponding observed frequencies. A numeral is placed to the right of each of these lines. Each numeral is the sum of the observed frequencies for the corresponding group of four patterns. Each is a sub-sum and will be used in Table XIII.

$$\chi^2 = \frac{(339 - 312.6)^2}{312.6} + \frac{(100 - 126.4)^2}{126.4} + \frac{(247 - 273.4)^2}{273.4} + \frac{(137 - 110.6)^2}{110.6}$$
$$= (26.4)^2 \left(\frac{1}{312.6} + \frac{1}{126.4} + \frac{1}{273.4} + \frac{1}{110.6}\right)$$
$$= 696.96\,(.02381) = 165.946.$$

The number of *df* is 1. This follows because there are four cells and because three restrictions were placed in determining the theoretical frequencies. The three restrictions were the total number of subjects (=823) and two border sums, one for a row and one for a column.[21] The calculated value of χ^2 is highly significant and leads immediately to a rejection of the hypothesis that the conditions of guessing had no influence on the tendency to guess heads on the first guess.

Table XI may be regarded as a 32×2 contingency table, 32 "patterns" by 2 "conditions of guessing." In order to use the χ^2 test in evaluating the hypothesis that the patterns were independent of the conditions of guessing, the border sums would be used in calculating theoretical frequencies. For example, the two theoretical frequencies for the pattern *H H H H H* are given by the relations: $F_t = (54 \times 439)/823 = 28.8$; and $F_t = 54 - 28.8 = 25.2 = (54 \times 384)/823$. These two values of F_t are shown in the top row of the last two columns of Table XI. The theoretical frequencies for the other 31 patterns were obtained in a similar way and are listed in the table. These frequencies are included to emphasize the fact that the χ^2 test cannot be legitimately applied to this table as it stands. The reason is that 37 of the 64 theoretical frequencies are less than 10 (some of them very much less than 10) and cannot be depended upon to yield quantities distributed as χ^2. To make a legitimate application of χ^2 in this particular case, it would be necessary to increase the number of subjects to a point where the smallest of the theoretical frequencies was close to 10.

It must now be decided whether or not the observed frequencies in Table XI can be combined so as to permit the use of the χ^2 test in evaluating the hypothesis of non-relationship between the patterning of the guesses and the conditions of guessing. The frequencies have already been combined in a gross way to yield the observed frequencies in Table XII. This division was on the basis of the first guess. Divisions might be made on the basis of the first two guesses or the first three guesses or the first four guesses. It is not possible to use the patterning on all five guesses because the theoretical frequencies become too small, as already seen. It turns out that a division on the basis of the first four and four in the other will be included. A total of 2,274 eighth-grade pupils, enrolled in 91 different schools, took an English Correctness Test. A summary of the scores obtained by these pupils has been taken from a report by Lindquist (**23**). The scores were divided into five categories and symbolized by numbers from 1 to 5, as seen in Table XIV. The schools were divided into four enrollment groups, labeled A to D in the table. The observed frequencies in the 20 cells of the table range from 3 to 342. These frequencies would obviously have been different if the enrollment groups had been differently established, and if the scores had been divided into different categories. The enrollment grouping of the schools was that commonly used in the Iowa Every-Pupil Testing Program.[22] The division of the scores was made by starting at the bottom and "stepping off" successive standard deviation "distances" (approximately). The distribution of scores was positively skewed; so it was necessary to combine the two upper score categories to provide satisfactorily large frequencies in the top row of cells.

---

[21] The sums 439 and 237 (or 384 and 237, or 384 and 586) could have been used instead of 586 and 439. The values of F_t for all but one cell may be obtained by subtraction.

The theoretical frequencies for Table XIV were computed in the usual way when the hypothesis of independence is under test. The computations required the use of three column sums and four row sums, as well as the value of $N(=2274)$. This made a total of 8 restrictions; so the number of $df=20-8=12=(5-1)(4-1)$. The value of χ^2 was computed as follows:

$$\chi^2=\frac{(36-23.5)^2}{23.5}+\frac{(76-60.2)^2}{60.2}+\cdots+\frac{(88-83.5)^2}{83.5}+\frac{(75-47.3)^2}{47.3}=99.038.$$

With $df=12$, this value is highly significant and leads to a rejection of the hypothesis that the scores obtained on the test were independent of school size.

A comment should be made concerning the theoretical frequency in the upper right-hand cell of the table. Its value is 8.9. Ordinarily, a value this small should not be used in obtaining an estimate of χ^2. In this case, however, the other 19 values of F_t are satisfactorily large, and the inclusion of one theoretical frequency that is less than 10 is permissible since an error in a single category will have slight effect on the resulting value of χ^2. The obtained value of χ^2 is so large that it makes no difference whether or not the small theoretical frequency is included in the calculations. It is only in situations of this general kind that one or two small theoretical frequencies may be retained. When the number of df is less than 4 or 5, and especially when $df=1$, the use of theoretical frequencies of less than 10 should be strictly avoided.

Use of the Chi-square Test with too Small Theoretical Frequencies

The studies of Lewis and Franklin (**21**) *and Lewis* (**20**). The commonest weakness in applications of the χ^2 test to contingency tables is the use of extremely small theoretical frequencies. This weakness is clearly present in most of the applications made in a paper by Lewis and Franklin (**21**). The paper is concerned with the Zeigarnik effect (that is, with the relative amounts of recall of interrupted and completed tasks). In one experiment, 12 subjects were each presented with 18 problems, 9 of which were interrupted by the experimenter, the other 9 being completed without interruption. The ratio (RI/RC) of the number of interrupted tasks recalled to the number of completed tasks recalled is given for each subject in the top row of Table XV. In a previous study, Lewis (20) had employed a cooperative work situation in which a co-worker completed the tasks on which the subject was interrupted. The conditions were otherwise the same as in the later experiment by her and Franklin. Fourteen subjects were used. The *RI/RC* ratio for each subject is shown in the second row of Table XV. The median ratio for Group I was .67. The ratios for the two groups were divided (dichotomized) at this point. The 2×2 contingency table shown as Table XVI was the result. The hypothesis to be tested was that the conditions of the two experiments had no differential effects on the recall of interrupted and completed tasks. The theoretical frequencies are shown in parentheses in the table. The investigators calculated a value of χ^2 in the usual way. Finding it to be approximately 3.8 and to fall near the 5% level of confidence, they were inclined to reject the hypothesis.

---

[22] The enrollment categories were: A, greater than 400; B, 126–400; C, 66–125; D, less than 66.

TABLE XV

DATA UPON THE "ZEIGARNIK EFFECT" PRESENTED BY LEWIS (20), AND LEWIS AND FRANKLIN (21)

Ratio of Recall Scores	*Subjects*													
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
RI/RC—Group I (Indiv. Work Condition)	1.00	.80	.80	.80	.75	.71	.63	.57	.57	.44	.40	.25		
RI/RC—Group II (Coop. Work Condition)	1.67	1.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.80	.75	.60	.14

Entirely aside from any of the conclusions reached by Lewis and Franklin, it must be firmly stated that all four of the theoretical frequencies in Table XVI, and especially the two that are less than 5, are too small to warrant an application of the χ^2 test. Furthermore, the other applications made in their paper, with one or two possible exceptions, involved theoretical frequencies that should be avoided (theoretical frequencies, for example, of 4.4, 6.0, 7.0, 7.6, etc.).

TABLE XVI

CONTINGENCY TABLE BASED UPON DATA IN TABLE XV

RI/RC	*Group*		
	I	II	
Greater than 0.67	6 (8.3)	12 (9.7)	18
Less than 0.67	6 (3.7)	2 (4.3)	8
	12	14	26

Kuenne's study of transposition behavior. An application, which is correct in principle but which must be regarded somewhat unfavorably because of the size of the F_t's, comes from a paper by Kuenne (18). Kuenne made a study of transposition behavior in four groups of young children, the ages for the groups being 3, 4, 5, and 6 years. Some of the children displayed size transposition, others of them did not. Kuenne realized that, because of the small number of cases in each age group, she could not apply the χ^2 test to the data for the four groups considered separately. Consequently, she combined the results for ages 3 and 4, and those for ages 5 and 6. The children were divided into two categories—those who did and those who did not meet the transposition criterion. The resulting 2×2 contingency table is shown as Table XVII. The hypothesis to be tested is that the occurrence of transposition behavior was independent of age. Theoretical frequencies were determined by using two of the border sums and the value of N. It will be seen that two of these frequencies are less than 10. Because they are fairly close to 10, many investigators would proceed as Kuenne did, and make a χ^2 test of the hypothesis of independence. In fact, there are textbooks in statistics which place 5 as the minimum value for theoretical frequencies. A value of 5 is believed to be too low. In any event, it is the smallest value that should be used even when there are several other theoretical frequencies that are far greater than 10.

TABLE XVII

CONTINGENCY TABLE BASED UPON KUENNE'S (18) TRANSPOSITION DATA

Age-Groups	*Transposition*	*Non-Transposition*	*Total*
3–4 years	3 (9.4)	15 (8.6)	18
5–6 years	20 (13.6)	6 (12.4)	26
	23	21	44

The value of χ^2 computed from Table XVII is 15.434. With 1 *df*, this value falls close to the 0.1% level. It is only because the value is so large that confidence can be placed in the conclusion that transposition behavior was related to age. In view of the smallness of all four of the theoretical frequencies, very great doubt would have remained if the χ^2 value had fallen at a border-line level of confidence. Whenever small theoretical frequencies enter into calculations of χ^2, the experimenter has no sound basis either for accepting or rejecting a hypothesis except when the value is quite extreme.

Yates' correction for continuity. No mention has as yet been made of a correction proposed by Yates (30) which reduces the value of χ^2 to compensate for errors which may arise as a result of one of the approximations made in deriving the formula for the χ^2 distribution. It will be recalled that three approximations are made in this derivation. One of the three involves the substitution of an integral for a summation of discrete quantities. This approximation introduces an error (an error of overestimation) that is of consequence when values of F_t are small. The correction is justified only when the number of *df* is 1.[23] It provides for the reduction of all differences between observed and theoretical frequencies by 0.5. For example, all of the differences between observed and theoretical frequencies in Table XVII are 6.4. These are reduced to 5.9 if Yates' correction is applied. The calculation, using the correction, would be as follows:

$$\chi^2 = (5.9)^2\left(\frac{1}{9.4}+\frac{1}{8.6}+\frac{1}{13.6}+\frac{1}{12.4}\right) = (34.81)(.3768) = 13.1.$$

With 1 *df*, this value is still highly significant and leads to a rejection of the hypothesis. But the use of Yates' correction does not remove the objection to theoretical frequencies that are less than 10.

Questionable or Incorrect Categorizing

Lewis' (20) *study of recall of interrupted and completed tasks.* Another weakness which is sometimes present in applications of the χ^2 test to contingency tables is that the categorizing is done on either a questionable or a clearly incorrect basis. An illustration of incorrect categorizing is found in the paper by Lewis (20) discussed above. The *RI/RC* ratios for 14 subjects were obtained in a "cooperative work experiment." The ratios are the ones given in the second row of Table XV. On the assumption that the recall of interrupted and completed tasks should have been the same, Lewis writes: ". . . we should have an equal number of ratios above 1.00 and below 1.00. . . . The expected distribution of ratios should, therefore, be 7 below 1.00 and 7 at 1.00 or above. The obtained distribution of ratios is 4 below 1.00 and 10 at 1.00 or above." The categorizing is plainly wrong; there is no more reason for placing an obtained ratio of 1.00 in the upper category than for placing it in the lower category. A better procedure would have been to divide the 8 ratios of 1.00 equally between the two categories. This flaw in Lewis' division of the ratios is remindful of the belief of many graduate students in psychology that it is quite permissible to set up several different sets of dichotomy lines, compute a value of χ^2 for each set, and finally select the dichotomies that yield a χ^2 value to support the experimenter's own point of view. In any investigation where the χ^2 test is to be applied, the categories must be established in a logically defensible and reliable manner—before the data are collected, if possible.

Anastasi and Foley's (1) *study of drawings of normal and abnormal subjects.* The whole problem of categorizing may be brought clearly before the reader by taking an illustrative case from a study by Anastasi and Foley (1). These two investigators had each of 340 normal subjects and 340 abnormal subjects draw a picture which depicted danger. The pictures were then divided into the 20 subject-matter categories listed in Table XVIII. The application of the χ^2 test yields a value of 99.603 which, with 19 *df*, falls far beyond the 0.1% level of confidence. This leads to a rejection of the hypothesis that the subject matter of the drawings was independent of the two "kinds" of subjects.

---

[23] The correction should not be made if several values of χ^2 are to be summated. The additive principle does not apply to corrected values.

TABLE XVIII

DATA FROM ANASTASI AND FOLEY'S (1) STUDY OF DRAWINGS OF NORMAL AND ABNORMAL SUBJECTS

Subject-Matter Categories	F_o (*Abnormal*)	F_o (*Normal*)	F_t	χ^2
1. Traffic	105	124	114.5	1.576
2. Conventional sign or signal	36	37	36.5	.014
3. Skating, ice	5	10	7.5	1.667
4. Falling	21	17	19.0	.421
5. Drowning, sinking, flood	8	10	9.0	.222
6. Falling objects, explosion	6	6	6.0	.000
7. Arms and explosives	10	22	16.0	4.500
8. War	3	7	5.0	1.600
9. Fire	26	37	31.5	1.921
10. Lightning, electricity	6	15	10.5	3.857
11. Animals	3	16	9.5	8.895
12. Abstract or conventionalized symbolism	3	8	5.5	2.273
13. Fantastic compositions	26	0	13.0	26.000
14. Several discrete objects	9	2	5.5	4.454
15. Scribbling or scrawl	6	0	3.0	6.000
16. Writing only	12	1	6.5	9.308
17. Miscellaneous	23	26	24.5	.184
18. Recognizable object not representing danger	8	1	4.5	5.444
19. Refusal to draw	14	1	7.5	11.267
20. No data for other reasons	10	0	5.0	10.000
	340	340		99.603

Let it again be emphasized that the criticism here, as elsewhere in the paper, is not directed at any of the conclusions reached by the investigators. But the illustration provides a very satisfactory basis for discussing the fundamental problem of categorizing. In the published article, the principles adopted in classifying the pictures are not explicitly stated. Furthermore, evidence is not presented regarding the reliability of the categories. Two generalizations may be offered. The first is that, *whenever possible, categories for frequency data should be established on the basis of completely external criteria* (for example, criteria that have been used or proposed by some other investigator) and should be set up independently of the data under study. Such a procedure frees a person from any charge of bias and guards against tendencies to juggle data. A second generalization is that *information on the reliability of categories should be offered*, and this is the case whether or not the categories have stemmed from an independent source.

A study of Table XVIII shows rather quickly that the value of χ^2 for the drawings depicting danger would have been quite different if the categorizing had been different. For example, the amount 26.000 was contributed to the value of χ^2 by the frequencies of category 13 alone. As seen, this is the category "fantastic compositions." The decision which established this category and the judgments which placed 26 of the drawings of the abnormal subjects in the category and none of the drawings of the normal subjects in it, should have been explicitly justified and a precise statement concerning reliability should have been included. The discrepancy between the frequencies in category 13 (along with the discrepancies in such categories as "refusal to draw" and "no data for other reasons") required that there be discrepancies in one or more of the other categories. Unreliability at one point in a multicelled table automatically produces unreliability elsewhere.

It is well to emphasize, by reiteration, that when the χ^2 test is to be applied to a collection of data, the categories should be established independently of the data and, once established, should never be modified on the basis of the way the data happen to fall. Categories should usually, if not always, be established before the data have been scrutinized.

APPLICATIONS: III. THE GOODNESS OF FIT OF FUNCTIONS IN WHICH FREQUENCY IS THE DEPENDENT VARIABLE

The frequency (or relative frequency) of occurrence of a response is sometimes used as the dependent variable in psychological experiments. For example, in psychophysical investigations based on the method of constant stimuli, the number or proportion of judgments in a given direction serves as the dependent variable, while in studies of the conditioned response, the frequency of occurrence of the CR is often taken as the dependent variable. If a mathematical function is fitted to data of this type, it is sometimes possible to apply the χ^2 test in evaluating the goodness of the fit. However, care must be taken to insure a correct application.

Goodness of fit of the phi-gamma function. It is a fairly common practice among psychophysicists to apply the χ^2 test to the differences between observed and theoretical proportions in order to evaluate the phi-gamma hypothesis—the hypothesis that the observed proportions can be represented satisfactorily by the phi-gamma function.[24] The usual procedure in making the test will be illustrated with some weight-lifting data.

As part of a laboratory exercise, a graduate student in an advanced experimental course made 100 judgments on each of nine pairs of weights. The weights, in grams, are listed in the X column of Table XIX. The method of constant stimuli was employed. The 100-gram weight was the standard and was paired 100 times, not only with itself, but with *each* of the other eight ("variable") weights. The standard weight was always lifted first, the variable weight second. Each of the 900 judgments was made in terms of the question: Which weight is the heavier, the first or the second? The second column of the

---

[24] The use of the phi-gamma function in psychophysical research is explained by Guilford (**13**, Chap. VI).

TABLE XIX

TEST OF GOODNESS OF FIT OF THE PHI-GAMMA FUNCTION TO WEIGHT-LIFTING DATA

X	p_o	γ	$X\gamma$	X^2	γ_t	p_t	(A)		(B)		
							$(p_o-p_t)^2$	$\frac{(p_o-p_t)^2}{p_tq_t}$	p_o	p_t	$n_k\frac{(p_o-p_t)^2}{p_tq_t}$
104	.97	1.3299	138.31	10,816	1.246	.960	.000100	.002600	.935	.934	.0032
102	.90	0.9062	92.43	10,404	.942	.908	.000064	.000768			
100	.79	0.5702	57.02	10,000	.638	.816	.000676	.004516	.79	.816	.4516
98	.65	0.2724	26.70	9,604	.334	.681	.000961	.004421	.65	.681	.4421
96	.53	0.0532	5.11	9,216	.030	.517	.000169	.000676	.53	.517	.0676
94	.37	−0.2347	− 22.06	8,836	− .274	.350	.000400	.001760	.37	.350	.1760
92	.20	−0.5951	− 54.75	8,464	− .578	.208	.000064	.000384	.20	.208	.0384
90	.13	−0.7965	− 71.68	8,100	− .882	.107	.000529	.005539	.085	.077	.1804
88	.04	−1.2379	−108.94	7,744	−1.186	.047	.000049	.001103			
$\sum$ 864		0.2677	62.14	83,184				.021767			1.3593

table, labelled p_o, gives the proportion of times (in 100) that each weight was judged heavier than the standard weight.[25]

A phi-gamma function may be fitted to the proportions in Table XIX, and the χ^2 test may then be applied (with a reservation to be specified later) to evaluate the goodness of fit. The phi-gamma function is basically the same as the equation for the normal ogive, which may be written

$$p = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}\, dx, \qquad [13]$$

with p standing for proportion of area under the normal curve from $-\infty$ to x. As usual, x represents a deviation of the variate X from the mean of X. The substitution in equation [13] of the symbol h for the quantity $1/\sigma\sqrt{2}$ yields

$$p = \int_{-\infty}^{x} \frac{h}{\sqrt{\pi}} e^{-h^2x^2}\, dx. \qquad [14]$$

In this expression, $x = X - L$, where L is the value of X corresponding to a proportion of .50. The constant h is an index of the steepness or "precision" of the ogive curve. The phi-gamma function, written directly from [14], takes the form

$$p = \int_{-\infty}^{\gamma} \frac{1}{\sqrt{\pi}} e^{-\gamma^2}\, d\gamma \qquad [15]$$

where γ has been substituted for hx.

In fitting the phi-gamma function to a set of observed proportions, it is necessary to obtain estimates of the constants h and L. Since $\gamma = hx$ and $x = X - L$, it follows that

$$\gamma = hX - hL. \qquad [16]$$

This equation is linear in γ and X. The principle of least-squares is invoked in deriving the following two equations with which estimates of h and L may be computed:

$$h = \frac{N\sum X\gamma - \sum\gamma\sum X}{N\sum X^2 - (\sum X)^2}; \qquad [17$$

$$L = \frac{\sum Xh - \sum\gamma}{Nh}. \qquad 18$$

In the derivation of these equations, the quantity that is minimized is

$$\sum(\gamma_o - \gamma_t)^2 = \sum(\gamma - hX + hL)^2.$$

Before equations [17] and [18] can be solved, values of γ must be obtained. They may be found in a table of the phi-gamma function (that is, in a table based on solutions of equation [15] for values of p).[26] The table is entered with values of p_o.

---

[25] The reason that the value of p_o for the 100-gram weight is significantly greater than .50 is that a negative time error was operating. This means that there was a consistent tendency for the subject to over-estimate the weight of the second member of a pair. The tendency was present in all comparisons, but was most immediately apparent from the value of p_o when the standard weight was compared with itself.

[26] Tables of the phi-gamma function (which give paired values of p and γ) are not usually included in statistics books published in the United States. However, it is easy

Values of γ corresponding to the nine values of p_o in Table XIX are given in the third column of the table. The fourth column gives values of $X\gamma$, the fifth column values of X^2. The four sums needed for solving equations [17] and [18] appear at the bottom of the table. Estimates of constants h and L are computed as follows:

$$h = \frac{9(62.14) - (.2677)(864)}{9(83,184) - (864)^2} = 0.152;$$

$$L = \frac{(864)(.152) - (.2677)}{9(.152)} = 95.8.$$

The insertion of these estimates into equation [16] gives

$$\gamma_t = 0.152X - 14.562.$$

This equation was used to calculate the values of γ_t shown in Table XIX. The corresponding values of p_t were read from a table of the phi-gamma function.

The χ^2 test may be applied (with a reservation) to test the goodness of fit, that is, to test the hypothesis that the difference between the observed and theoretical proportions arose from sampling fluctuations and, consequently, that the observed proportions may be satisfactorily represented by the phi-gamma function. The procedure for calculating the value of χ^2 based on all nine theoretical proportions is illustrated in the two columns of Table XIX denoted by (*A*). Values of $(p_o - p_t)^2$ were first obtained and were then divided by the product p_tq_t. This single division by p_tq_t took account of differences between p_o and p_t as well as differences between q_o and q_t (where, as usual, $q_o = 1 - p_o$ and $q_t = 1 - p_t$). It can easily be shown that

$$\frac{(p_o - p_t)^2}{p_t} + \frac{(q_o - q_t)^2}{q_t} = \frac{(p_o - p_t)^2}{p_tq_t}.$$

Emphasis is here given again to the fact that in the calculation of χ^2, account must be taken of the frequency of occurrence *and* the frequency of non-occurrence of an event.

The sum of the second column under (*A*) is .021767. This is not the value of χ^2. However, it becomes the value of χ^2 after it has been weighted by 100, the number of judgments upon which each value of p_o was based. In other words,

$$\chi^2 = n_k \sum \frac{(p_o - p_t)^2}{p_tq_t} = 100\ (.021767) = 2.1767$$

where n_k is the number of judgments per proportion. The number of *df* is $7 = 9 - 2$. It is two less than the number of rows in the table.[27] The computed

---

to obtain values of γ indirectly from an ordinary table of the probability integral. Since $z = x/\sigma$; $\gamma = hx$, and $h = 1/\sigma\sqrt{2}$, it follows that $\gamma = .7071z$ and $z = 1.4142\gamma$.

[27] The number of degrees of freedom for a table such as XIX is readily determined by starting with the number of rows. Each row contributes 2 *df* before any restrictions are imposed. (This is because each row actually covers two pairs of proportions. The pair which is omitted in the formula is q_o and q_t.) The restriction that $(p_t + q_t) = 1.00$ is imposed on *each* row; so 1 *df* is lost for each. Further, 1 *df* is lost for each constant estimated

value of χ^2 falls near the 95% level of confidence; so the phi-gamma hypothesis cannot be rejected.

The applicability of the last statement rests upon the validity of applying the χ^2 test in the manner illustrated. There are two weaknesses in the application, one relating to the size of two of the theoretical frequencies and the other relating to the restrictions imposed during curve fitting. Before the frequencies are considered, a few comments will be made on the nature of the restrictions that were imposed, although it must be said that a meaningful explanation of restrictions is mathematically rather complex and lies beyond the scope of this paper. Whenever a curve is fitted to a set of empirical data, there are usually several acceptable methods of determining values for the constants to be inserted in the equation. It is relatively easy to show that some of these methods lead to values of "χ^2," as computed by the familiar formula, which tend to be consistently larger or smaller than the values obtained when other curve fitting methods are employed. Therefore, when a value of χ^2 is computed to test the goodness of fit of such functions as the phi-gamma function, the method used to estimate the constants becomes a matter of importance. The computed quantity is distributed as χ^2 only when the restrictions imposed during curve fitting are both *linear* and *homogeneous*. (The meaning of these terms is explained more fully in a later section.) In many instances of curve fitting, and the present one is a case in point, the restrictions that are imposed do not have the requisite properties. When this is true, the investigator must be especially cautious in interpreting his results.

Some consideration must now be given to the theoretical frequencies represented by the theoretical proportions in Table XIX. The proportion .960 in the first row corresponds to two theoretical frequencies: 96.0 and 4.0. These values are estimates of the number of judgments "heavier" and the number of judgments "lighter" that the subject should have made when comparing the 104-gram weight with the standard weight. The proportion .047 in the bottom row also corresponds to theoretical frequencies of 4.7 and 95.3. These frequencies indicate that, according to the fitted function, the subject should have said "heavier" 4.7 times and "lighter" 95.3 times when comparing the 88-gram weight with the standard. The two frequencies of 4.0 and 4.7 are too small to yield quantities distributed as χ^2 and should not have been used in the calculations. They were included in the illustration as a means of contrasting the correct with the incorrect procedure. The two extreme proportions should have been combined with adjacent proportions. Table XIX shows the results after the combining has been done. New values of p_o, based on .97 and .90, and also on .13 and .04, are shown in the column labeled p_o under the general heading (*B*). Corresponding theoretical proportions are given in the column labeled p_t. The two combined values of p_o (.935 and .085) are averages based on 200 judgments each and the quantity $(p_o - p_t)^2/p_tq_t$ calculated for each must be weighted by 200 instead of 100. Because of this differential weighting, it is better to multiply by n_k, as shown in the last column of the table, before the

---

from the data during the curve fitting process. Since two constants (*h* and *L*) must be estimated, the number of *df* is 2 less than the number of rows (or the number of pairs of p_o and p_t) that enter into the calculations.

The study was designed to evaluate the theory of the neural quantum in auditory discrimination. The theory required, among other things, that the observed proportions for a subject could not be adequately represented by the phi-gamma function.[28] Consequently, phi-gamma functions were fitted to each of the six sets of proportions in Table XX. All proportions greater than .97 and less than .03 were omitted during the curve fitting process. A few of these proportions later played a part in the calculations of χ^2. This is one of the pitfalls to be carefully avoided. Another important aspect of the curve-fitting procedure was that the Müller-Urban weights were utilized. The nature of these weights and the detailed procedures for treating weighted data are adequately explained by Guilford (**13**) and need not be discussed here. It is enough to say that the weights are designed to diminish the influence of extreme proportions in the computations of the constants *h* and *L*.

TABLE XXI

VALUES OF *h* AND *L* SECURED BY STEVENS, MORGAN AND VOLKMANN (**32**)

Subject	*h*	*L*
JV	0.8520	3.9120
GS	1.2098	3.0525
RR	0.8476	4.1051
JM	1.0640	2.9001
DM	0.8754	3.5885
MJ	2.0432	1.6194

The values of *h* and *L* for the six fitted functions are shown in Table XXI. These constants were used in the manner illustrated in Table XIX to obtain theoretical proportions for each set of observed proportions. The theoretical proportions used by the investigators in calculating values of χ^2 are shown in Table XX. The six calculated values of χ^2 are given in the first row of Table

---

[28] The theory also required that the proportions for the several subjects could be fitted satisfactorily with straight-line functions. The results for the straight-line fits will be omitted here. It is obvious, however, that a straight line can be fitted to each set of proportions in Table XX as a means of obtaining theoretical (calculated) proportions. The χ^2 test could then be applied to the differences between the observed and theoretical proportions, in a manner exactly analogous to that employed when theoretical proportions are obtained with phi-gamma functions. Whenever a comparison is to be made between the goodness of fit of two different functions, the same quantity should be minimized in the process of obtaining the constants for the functions. In their study, Stevens, Morgan, and Volkmann minimized the sum of squared differences between observed and theoretical values of gamma in fitting phi-gamma functions but minimized the sum of squared differences between observed and theoretical proportions in fitting straight-line functions. According to a theorem given by Cramér (5, p. 426 ff.), the values of χ^2 obtained with the phi-gamma functions would be expected to be larger than those obtained with the straight lines.

XXII, along with the number of *df* for each. (The other two rows of values in this table will be explained later.) Only one of the values in the first row (the value for subject *RR*) is sufficiently large to justify the rejection of the phi-gamma hypothesis at better than the 5% level of confidence. However, the six separate values of χ^2 are completely independent estimates and, therefore, may be summated to provide a single composite value of χ^2, the number of *df* for which is the sum of the six separate *df*'s. The composite value appears in the

TABLE XXII

VALUES OF CHI-SQUARE BASED UPON THE DATA OF STEVENS, MORGAN AND VOLKMANN (32)

	Subjects						
	JV	*GS*	*RR*	*JM*	*DM*	*MJ*	*Composite*
χ^2	7.64	4.02	10.35	2.97	5.90	2.92	33.80
df	5	3	4	3	3	3	21
χ^2	4.94	0.13	4.06	2.97	5.90	1.02	19.02
df	4	2	3	3	3	2	17
χ^2	1.70	0.13	3.14	2.87	2.90	1.02	11.76
df	2	1	2	1	2	2	10

TABLE XXIII

SHOWING THE EFFECTS UPON CHI-SQUARE VALUES OF USING OBSERVED PROPORTIONS WHICH WERE EXCLUDED IN THE DERIVATION OF THEORETICAL PROPORTIONS (DATA FROM TABLE XX)

				(*A*)	(*B*)			(*C*)
X	p_o	p_t	$(p_o-p_t)^2$	$\frac{(p_o-p_t)^2}{p_tq_t}$	$\frac{(p_o-p_t)^2}{p_tq_t}$	p_o	p_t	$n_k\frac{(p_o-p_t)^2}{p_tq_t}$
3.0	(.02)	.093	.005329	.0629	—	—	—	—
3.5	.21	.234	.000576	.0032	.0032	.21	.234	.32
4.0	.52	.450	.004900	.0201	.0201	.52	.450	2.01
4.5	.64	.682	.001764	.0081	.0081	.64	.682	.81
5.0	.84	.858	.000324	.0027	.0027	.905	.905	.00
5.5	.97	.953	.000289	.0065	.0065			
				.1035	.0406			3.14

From Column A: $\chi^2 = 100\ (.1035) = 10.35$, with 4 *df*
From Column B: $\chi^2 = 100\ (.0406) = 4.06$, with 3 *df*
From Column C: $\chi^2 = 3.14$, with 2 *df*.

last column of the table and is 33.80, with 21 *df*. This value falls at about the 3% level of confidence and indicates that the fits of the phi-gamma curves were generally poor.

A serious flaw in this application was that four of the extreme observed proportions, no one of which played any part in determining constants h and L, were included during the calculations of the values of χ^2. A specific example will clarify the point. The observed and theoretical proportions for subject *RR* are listed in the second and third columns of Table XXIII. The observed proportion .02, shown in parentheses, was purposely excluded by the investigators when making least-squares solutions for h and L. Once obtained, these constants were crucial in determining the theoretical proportions. The investigators decided to include all observed proportions whose corresponding theoretical proportions were neither greater than .97 nor less than .03. This meant including the observed proportion .02 in the calculations for Table XXIII. It also meant the inclusion of observed proportions 1.00, .00, and .99 in the calculations for subjects JV, GS, and MJ, respectively.

As shown in Table XXIII, the estimated value of χ^2 was 10.35 when $p_o = .02$ and $p_t = .093$ were included in the calculation. If this pair of proportions is not included, the estimated value is 4.06. The number of *df* for the first estimate is 4, for the second, 3. There can be no doubt that the second estimate is the better of the two. In the calculation of χ^2, it is incorrect to use observed proportions which have not been allowed to influence the magnitude of the theoretical proportions.

The values of χ^2 shown in the second row of Table XXII were obtained by excluding the observed proportions which played no role in the curve-fitting process. There are still six independent values, each with its own number of *df*. It is entirely legitimate to add these values to obtain the composite value of 19.02 (with 17 *df*) given in the last column of the table. This new composite value falls at about the 30% level and provides no basis for rejecting the phi-gamma hypothesis for the six sets of proportions.

Still another flaw in the calculations of the values in Table XXII was that theoretical proportions representing theoretical frequencies of less than 10 were not first combined with adjacent theoretical proportions. For example, the theoretical proportion .953 (in Table XXIII) represents theoretical frequencies of 95.3 and 4.7. A theoretical frequency of 4.7 is too small to yield a quantity distributed as χ^2. Therefore, theoretical proportion .953 should have been combined, as shown in Table XXIII, with the theoretical proportion .858 to give $p_o = .905$. The modified value of χ^2 is 3.14, with 2 *df*. This value appears for subject RR in the third row of Table XXII. Similarly modified values for three other subjects are included in the same row. Again there are six independent estimates of χ^2, and these may be added to obtain the single composite value of 11.76, listed in the last column of the table. With 10 *df*, this estimate falls around the 30% level.

It will be realized that the elimination of the small and large proportions makes impossible a really critical evaluation of the phi-gamma hypothesis through the use of the χ^2 test. But this is not a fault of χ^2; it is a weakness in the experimental data. Whenever χ^2 is to be employed, the experimenter must take precautions to insure theoretical frequencies of adequate size.

The principal pitfalls in the use of χ^2 with proportions have been designated.

Two others should be mentioned—the tendency to divide $(p_o - p_t)^2$ by p_t alone instead of by the product $p_t q_t$ without including frequencies of non-occurrence, and the tendency to neglect to weight by the number of judgments upon which the proportions are based.

An application of χ^2 to percentages (Grant and Norris, **10**). A recent application of the χ^2 test to percentages (or relative frequencies) is taken from a paper by Grant and Norris (**10**). These investigators were concerned with the influence of different amounts of dark-adaptation on the sensitization of the beta-response of the human eyelid to light.[29] One of their measures of degree of sensitization was the frequency of occurrence of the response. A subject looked straight ahead into a small box-like enclosure which was painted flat black inside and out. The stimulus was a small circle of light emitted from a circular milk-glass plate, 10 cm. in diameter, located at the back of the enclosure. When illuminated, the plate had a surface brightness of 241 millilamberts. The duration of the stimulus was about 750 milliseconds.

Four experimental conditions were employed, all subjects serving in each condition. The conditions differed in the amount of dark-adaptation present in the subjects. The amount of dark-adaptation depended upon the total length of time the subjects spent in darkness. The measure of amount of adaptation was the product *It*, where *I* was the surface brightness of the stimulus plate and *t* was the number of seconds spent in darkness. The *It* products for the four conditions were 28,920; 187,980; 347,040; and 506,100. The corresponding values of *t* were 120, 780, 1414, and 2100 seconds.

TABLE XXIV

DATA FROM THE EXPERIMENT OF GRANT AND NORRIS (10)

	Conditions			
	1	2	3	4
P_o	13.2	30.0	45.6	51.5
P_t	11.05	36.10	44.30	49.35

Thirty-three subjects participated in the experiment. A subject was first dark-adapted for 120 seconds. The stimulus light was then presented four times, with a "control" trial between the first two and last two presentations. The control trial served as a check on possible conditioning. The four stimulus trials were separated by dark intervals of 35 sec. on the average. The first four stimulus trials (coming after 120 secs. in darkness) were the trials for condition 1. Condition 2, in which the stimulus light was again presented four times along with a control trial, came after the subject had spent a total of 780 secs. in darkness. Conditions 3 and 4, with the stimuli presented in the same general fashion, came after 1414 and 2100 secs. in darkness.

It should be noted that each of the thirty-three subjects was given four stimulus trials in each of the four dark-adapted conditions. A count was made of the number of beta-responses occurring in each subject in each condition. The results were then combined to provide frequencies of occurrence of the response for each condition. These frequencies were used in computing percentages. The four observed percentages are shown in the first row of Table XXIV. They indicate the *relative* frequency of occurrence of the beta-response in each of the experimental conditions. In condition 1, for example, 13.2% of the 132 possible responses of the eyelid displayed the beta-response. The other three percentages may be similarly interpreted.

---

[29] The beta-response is one of two reflexes displayed by the eyelid when the eye is light-stimulated.

A logarithmic function was fitted to the data. The two variables were It, the amount of dark-adaptation, and P_o, the relative frequency of occurrence of the beta-response. The fitted function was: $P_t = 13.38 \log_e It - 126.39$.[30] Its solution for the empirical values of It yielded the values of P_t given in the bottom row of Table XXIV.

A value of χ^2 was calculated as follows:

$$\chi^2 = \frac{(13.2 - 11.05)^2}{11.05} + \frac{(30.0 - 36.10)^2}{36.10} + \cdots + \frac{(51.5 - 49.35)^2}{49.35} = 1.579.$$

Two constants were estimated from the data, leaving 2 *df*. With this number of *df*, the obtained value of χ^2 falls near the 40% level of confidence. This led the investigators to conclude that the data could be satisfactorily represented by a logarithmic function.

There are four mistakes in this application of the χ^2 test: two computational mistakes and two "theoretical" mistakes. The computational mistakes will be discussed first. The calculated value of χ^2 was not corrected to take account of the use of percentages instead of frequencies. It should have been multiplied by the ratio 132/100. The second computational mistake was a failure to take account of the frequency of non-occurrence of the beta-response in each of the four conditions. In condition 1, for example, the beta-response occurred 13.2% of the time and failed to occur 86.8% of the time. This latter percentage played no part in the calculation of χ^2. All four percentages of non-occurrence should have been employed. The value of χ^2 for Table XXIV, when correctly computed, is 3.081. The number of *df* is still 2.

The other two mistakes were more basic. The use of χ^2 to test goodness of fit was not warranted, for two reasons. In the first place, there were linkages within conditions. Each of the subjects was given four trials in each of the four conditions of dark-adaptation. The results for the subjects were pooled. There were undoubtedly individual differences in capacity to display the beta-response; so there must have been linkages within conditions. In applying the χ^2 test, the investigators assumed, in effect, that the 132 trials per condition were given to 132 instead of 33 subjects. The test was inapplicable because the assumption could not justifiably be made.

The second reason that the χ^2 test should not have been used arises from the lack of independence from condition to condition. As already stated, a fundamental requirement for the use of the χ^2 test is independence between individual measures. The differences between individual subjects that manifested themselves in any condition were certain to be maintained in the other conditions; so it cannot be assumed that the values of P_o in Table XXIV are unrelated.

---

[30] With common logarithms, the function becomes:

$$P_t = 30.81 \log_{10} It - 126.39.$$

APPLICATIONS OF THE χ^2 TEST TO NON-FREQUENCY DATA

Some investigators are prone to use a χ^2 test of goodness of fit whenever a set of observed and theoretical values of any kind is available for comparison. This mistake apparently grows from a misinterpretation of the well-known formula for computing χ^2:

$$\chi^2 = \sum \frac{(F_o - F_t)^2}{F_t}. \qquad [10]$$

Because this equation involves differences between observed and theoretical values, the conclusion is reached that the summation of the weighted squares of differences between observed and theoretical quantities yields a meaningful estimate of χ^2. Formulas superficially resembling equation [10] have been applied to non-frequency data, where theoretical values have been obtained from fitted curves.

Suppose that a study has been made of the amount of activity displayed by several groups of white rats deprived of food for differing lengths of time. Enough rats have been included in each deprivation group to provide means that are quite stable. These means, when plotted against time of food deprivation, show a systematic trend; so a curve is fitted to them. The equation for the curve permits the calculation of theoretical means which may be compared with the observed means. To test goodness of fit, differences between the observed and theoretical means are obtained. Each difference is squared and then divided by the appropriate theoretical mean. The sum of these weighted squared differences is taken as a meaningful estimate of χ^2.

Two of the fourteen papers (**4**, **16**) referred to in the opening paragraphs contain applications of this type. The fallaciousness of such applications becomes obvious when it is realized that values of χ^2 computed from non-frequency data vary in magnitude with the size of the units employed in measurement. Assume that two investigators have a common aim: To determine how the height of human males varies with age. They make measurements of the same individuals in various age groups, fit equations of the same form to their data, use these equations to compute theoretical values of height, and then obtain estimates of χ^2 in a manner similar to the one described above. One of the investigators has measured height in centimeters, the other in inches. Except for incidental discrepancies, the value of χ^2 calculated from the centimeter data will turn out to be 2.54 times the value calculated from the inch data.

The χ^2 test of linearity of regression. A χ^2 test is recommended in certain textbooks (**26**, p. 319; **12**, p. 237) as suitable for use, with non-frequency data, in evaluating linearity of regression. The formula, as usually presented, is as follows:

$$\chi^2 = \frac{\eta^2 - r^2}{1 - \eta^2}(N - k). \tag{19}$$

In this formula, η stands for correlation ratio; r for product-moment coefficient of correlation; N for the total number of measures; and k for the number of columns (groups) into which the measures have been divided. The number of *df* is $k-2$. The formula yields a variable, the distribution of which approximates the χ^2 distribution, under certain conditions. It can be used with some degree of confidence provided that N is quite large and k is quite small, and provided also that the measures from column to column are independent and homoscedatic as well as normally distributed.

An exact test of linearity of regression, which is applicable whenever equation [19] is, can be made by computing an F-ratio. The formula is

$$F = \frac{(\eta^2 - r^2)(N - k)}{(1 - \eta^2)(k - 2)}. \tag{20}$$

The *df*'s for this F are $(k-2)$ and $(N-k)$. The reason that the χ^2 test, as defined by equation [19], can be substituted under any circumstances for the F test, as defined by equation [20], is that the distribution of F approximates the distribution of χ^2 when one of the *df*'s for F is very small and the other is very large. The particular χ^2 distribution that is approximated is the one for the smaller *df* (that is, for $df = k-2$). To state the point in another way: The sampling distribution of estimates of F, obtained with equation [20], approximates the sampling distribution of estimates of χ^2 obtained with equation [19], provided that k is very small relative to N.

Inasmuch as the χ^2 test, as represented by equation [19], is inexact, while the F test, as represented by equation [20], is exact, nothing is gained by using the χ^2 test.[31]

SPECIAL PROBLEMS: I. INDETERMINATE THEORETICAL FREQUENCIES

It sometimes happens, despite superficial indications to the contrary, that meaningful theoretical frequencies cannot be determined for a set of observed frequencies. This situation commonly arises from a lack of independence between measures. Two illustrations will be presented to reveal some of the chief sources of difficulty.

First illustration. The first illustration is concerned with coin-guessing data. Two hundred and forty university students each made a guess of heads or tails on four successive tosses of a coin. They recorded their guesses on individual record sheets. They were not told in advance the number of tosses that would be made, nor were they told how the coin turned up on the four tosses until the record sheets had been collected. The turns, in order of occurrence, were

---

[31] A fuller explanation of equation [20], together with a detailed discussion of other F tests of goodness of fit, is given by Lewis (**19**).

H T T H. The succession of guesses of each student could easily be compared with this succession of turns and the number of correct guesses for each could be tabulated. As expected, the number of correct guesses ranged from none to 4. The results are summarized in the first two columns of Table XXV. As shown, 15 students made no correct guesses; 47 made one correct guess; 60 made two correct guesses, etc. With frequency distributions of this type, it is common practice to apply the chi-square test after a binomial distribution function has been employed to calculate the required theoretical frequencies. The function for the present case, modeled after equation [3], would have the form

TABLE XXV

ANALYSIS OF COIN-GUESSING DATA

Number Correct	F_o	$P(n)$	F_t	$\frac{(F_o - F_t)^2}{F_t}$	$P(n)'$	F_t'	$\frac{(F_o - F_t')^2}{F_t'}$
0	15	.0625	15	.000	.0490	11.76	.893
1	47	.2500	60	2.817	.1680	40.32	1.107
2	60	.3750	90	10.000	.3129	75.10	3.036
3	86	.2500	60	11.267	.3289	78.93	.633
4	32	.0625	15	19.267	.1412	33.89	.105
$\sum$	240	1.0000	240	43.351	1.0000	240.00	5.774

$$P(n) = \frac{4!}{n!(4-n)!}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^{4-n} = \frac{4!}{n!(4-n)!}\left(\frac{1}{2}\right)^4, \qquad [3']$$

and would yield theoretical probabilities of obtaining n correct guesses in 4.

Values of $P(n)$ calculated with equation [3′] are given in the third column of Table XXV with corresponding theoretical frequencies appearing in the fourth column. If these values of F_t were legitimate estimates, a value of χ^2 could be computed as illustrated in the fifth column of the table and used to test the hypothesis that the observed frequencies are distributed in accordance with equation [3′]. In effect, this would be testing the hypothesis that each guess of every student was a purely chance occurrence, completely independent of every other guess, the probability of a guess of heads always being .50. The number of *df* is 4. If the hypothesis were true, the obtained value of χ^2, 43.351, would not be expected to arise in random sampling once in a million times.

The use of a binomial distribution function in calculating theoretical probabilities cannot be justified in this case. There is good reason for believing that the guess of a student on any of the last three tosses was not independent of previous guesses. The binomial distribution function is applicable only when there is a sound basis for assuming that every event under consideration is completely independent of every other event.

Theoretical frequencies for the empirical data in Table XXV are indeterminate—except as they might be estimated from probabilities yielded by extraneous empirical data. When hypothetical probabilities do stem from other empirical results, the hypothesis that can be tested may be quite different from the one that would be tested if a binomial distribution function yielded the theoretical values. For example, the data of Table XXV may be compared with somewhat similar findings published several years ago by Goodfellow (9), who analyzed the patterning of the guesses on five successive tosses of a coin by a large number of radio listeners. These listeners participated in the "telepathic experiments" conducted in 1937–38 by the Zenith Foundation. The coin was tossed in the broadcasting studio in Chicago. A total of 5,687 members of the radio audience wrote down their guesses in order and mailed in their answer sheets. They were told in advance that they were to make five guesses, but were not told until two or three weeks after the broadcast how the coin had actually turned up each time. The conditions of guessing were not identical with those holding when the data in Table XXV were secured, but were similar enough to permit a comparison of the results.

Goodfellow (9, Table II) tabulated the results on the radio listeners in a way closely approximating that used in Table VIII. In fact, the only difference of any consequence is that he tabulated percentages instead of frequencies. The "correct" pattern of *H T T H* was checked against the 32 patterns in Goodfellow's table to obtain the proportions of individuals in the total group of 5,687 that hypothetically made 0, 1, 2, 3, and 4 correct guesses. These proportions are given in Table XXV, in the column headed F_t'. The resulting value of χ^2, as shown in the last column of the table, is 5.774 with $df = 4$. The hypothesis that may now be tested is that the patterning of the guesses by the 240 students was the same as the patterning of the first four of five guesses by the large group of Zenith Foundation listeners. The obtained value of χ^2 falls near the 20% level of confidence and provides no basis for rejecting this hypothesis. If the value of χ^2 *had been* large enough to justify a rejection of the hypothesis, it would not be possible to decide whether the patterning tendencies were basically different in the two groups or whether the differences in the conditions of guessing produced an apparent difference in patterning. Nevertheless, the test of the hypothesis, as stated, is an exact one. This is in sharp contrast to the indefiniteness which was present when a value of χ^2 was based on theoretical frequencies obtained with a binomial distribution function. Because of the strong likelihood of interdependence between the guesses of individual guessers, the highly significant value of χ^2 in the fifth column of Table XXV could be interpreted to mean that all guessers were biased, that some of the guessers were strongly biased while others were unbiased, that the probability of a guess of heads was not .50, that the probability of a guess of heads was .50 on some tosses but not on others, etc. The absence of independence and the consequent inability to obtain unequivocal theoretical (chance) frequencies made this application of the chi-square test a meaningless procedure.

Second illustration. Another illustration of the indeterminateness of theoretical frequencies comes from a paper by Seward, Dill and Holland (28). These investigators were concerned with an aspect of learning theory which need not be explained in order to describe their application of χ^2. The experimental procedure was relatively simple. A subject sat at a table facing a panel. On the table in front of him was a row of twelve push buttons. Ten colored cards were used as stimuli. They were exposed, one at a time, in a small rectangular window in the panel.

In the "learning" series, the colors (except blue-green) were exposed once each in a predetermined order and the subject pressed the buttons until one was found which turned on a light. Nine of the ten colors were each paired with one of the buttons. The blue-green color was associated with two buttons, the fifth and eighth. The ninth button was a blank. The blue-green color was presented twice in the "learning" series. When first presented, the light was connected with the fifth button for half of the subjects and with the eighth button for the other half. On the second presentation, the connections were reversed. Thus each subject had an opportunity to develop associative connections between blue-green and button 5, and also between blue-green and button 8.

In the "test" series, the ten colors were again presented in a predetermined order and each subject was given six chances to push the correct button for each color. In this series, blue-green was exposed last and the connection was such that the pressing of botton 5 *or* button 8 would turn on the light. The testing on any color was terminated as soon as the subject pushed the button that turned on the light or had pushed six buttons.

The investigators applied the chi-square test to the results on blue-green in an attempt to determine whether or not learning had occurred. Values of F_o and F_t, taken directly from their Table I (**28**, p. 231), are shown in the first two rows of Table XXVI. As seen, 24 of the 110 subjects pressed either button 5 or button 8 on the first trial. Twenty-five subjects pressed one of these two buttons on Trial 2, 22 on Trial 3, etc. Ten of the subjects failed to find either of the correct buttons in six trials.

TABLE XXVI

DATA FROM THE LEARNING EXPERIMENT OF SEWARD, DILL AND HOLLAND (**28**)

	Trials Required to Find Correct Button							*Total*
	1	*2*	*3*	*4*	*5*	*6*	*>6*	
F_o	24	25	22	21	4	4	10	110
F_t	18.3	16.7	15.0	13.3	11.7	10.0	25.0	110
p	2/12	2/11	2/10	2/9	2/8	2/7	?	
n	110	91.7	75.0	60.0	46.7	35.0	25.0	

The values of F_t in the second row of Table XXVI were obtained by assuming that the probabilities of success on the six trials were those given in the third row. In other words, p was assumed to be 2/12 on trial 1, 2/11 on trial 2, 2/10 on trial 3, etc. A value of χ^2 was computed as follows:

$$\chi^2 = \frac{(24 - 18.3)^2}{18.3} + \frac{(25 - 16.7)^2}{16.7} + \cdots + \frac{(10 - 25.0)^2}{25.0} = 31.3.$$

The only restriction imposed during the calculations was that $\sum F_o = \sum F_t = 110$. This left 6 degrees of freedom. With this number of *df*, the computed value of χ^2 is highly significant. It led the investigators to reject the hypothesis that the button pushing occurred by chance and to conclude that the subjects had, in fact, learned to associate blue-green with button 5 and/or button 8.

This application of the chi-square test was unwarranted because unequivocal theoretical frequencies for all trials, except possibly for trial 1, were indeterminate. The reason is that the events on trials 2 to 6 were related to previous events. The investigators seemed to be partially aware of this fact, as is shown by their choice of values of p. For example, consider the probability of succces on trial 4; it was put at 2/9. This probability arose from the assumption that each subject remembered and avoided the three buttons that had already been pushed, completely forgetting, meanwhile, all choices that had previously been made when other colors were presented. Just prior to the blue-green test series, the subjects had each pushed up to six buttons in relation to each of the other nine colors. It is inconceivable that the subjects could have remembered perfectly the buttons that had been "used up" either during the blue-green series or before on other series, unless they adopted "systems" of pressing—a system, for example, of pressing from either end toward the center, or from the center toward either end, or from the third button toward the right, or from the tenth button toward the left, etc. The adoption of any system or systems of pressing by any subject would mean that any hypothesis concerning chance proportions was open to rejection prior to the calculation of a value of χ^2.

When a decision concerning the legitimacy of theoretical probabilities is difficult to reach, it is sometimes helpful to see what happens if different probabilities are secured on the basis of new and not-unreasonable assumptions. It could be assumed, for example, that the 110 subjects used by Seward, Dill and Holland had, during the learning and early test series, discovered (on the average) that three of the twelve buttons (the two at the ends perhaps, and one in the middle) "belonged" to three colors other than blue-green, and thus by inference *did not belong* to blue-green. When blue-green was finally exposed, three of the buttons were already "used up" and were not likely to be pressed. The theoretical frequencies might be the ones given in the second row of Table XXVII. A value of χ^2 for this table is calculated as follows:

TABLE XXVII

ALTERNATIVE ANALYSIS OF THE DATA OF SEWARD, DILL AND HOLLAND (28)

	Trials Taken to Find Correct Button							*Total*
	1	*2*	*3*	*4*	*5*	*6*	*6*	
F_o	24	25	22	21	4	4	10	110
F_t	24.4	21.4	18.3	15.3	12.2	9.2	9.2	110
p	2/9	2/8	2/7	2/6	2/5	2/4	?	
n	110	85.6	64.2	45.9	30.6	18.4		

$$\chi^2 = \frac{(24 - 24.4)^2}{24.4} + \frac{(25 - 21.4)^2}{21.4} + \cdots + \frac{(10 - 9.2)^2}{9.2} = 12.005.$$

With 6 df, this value falls slightly below the 5% level of confidence. It rests on dubious assumptions and, regardless of its confidence level, provides no grounds for meaningful conclusions; but the fact that it differs so greatly from the χ^2

value of 31.3 computed for Table XXVI strongly suggests that unequivocal theoretical frequencies are indeterminate in both cases.

It is not possible to provide rules-of-thumb for deciding whether theoretical frequencies are calculable or incalculable in particular situations. Decisions must ordinarily be based on careful logical analysis. However, *it is usually true that theoretical frequencies are incalculable if the observed frequencies are in any way related, and also if mutually contradictory assumptions can be made, with about equal justification, concerning the likelihood of occurrence or non-occurrence of the events (responses) that yielded the observed frequencies.*

SPECIAL PROBLEMS: II. THE NATURE OF IMPOSED RESTRICTIONS

All restrictions that are imposed during the determination of theoretical frequencies should be both linear and homogeneous. This limitation is seldom mentioned in either theoretical or practical treatments of the chi-square test, and even when mentioned is usually left unexplained. The main reason for the omission is that anything beyond a very superficial explanation cannot be given in other than mathematical terms.[32] It follows that only a few very general ideas can profitably be included here.

It is probably quite obvious to most readers why restrictions must be imposed—why the sums of the observed and theoretical frequencies, for example, must always be equalized. The value of ΣF_o is fixed for any set of empirical data. The value of ΣF_t cannot "wander around any place" without at times yielding entirely impossible cell frequencies. Any hypothesis must be tied down at some point to the sample data. There should be as much freedom as possible (for example, as much freedom as possible for fluctuations in the individual cell frequencies) and yet there must be enough restrictions to bring the over-all values within the same general area.

The one restriction that must always hold may be symbolized as follows:

$$\sum F_o = \sum F_t, \qquad [21]$$

or

$$\sum(F_o - F_t) = 0. \qquad [21a]$$

---

[32] The excellent attempt at elementary explanation offered by Greenhood (11, Chap. 3) is about as non-mathematical as it could conceivably be made, and yet requires a considerable amount of mathematical sophistication.

This restriction is clearly linear.

Some of the other restrictions that must be imposed in familiar applications of the chi-square test may be shown to be linear—and to be homogeneous also. Suppose, for example, that a normal curve is to be fitted to an array of observed frequencies. It is not enough to impose the single restriction specified by equation [21]. The reason is that a multitude of different combinations of values of F_t, all arising from normal distributions, will each, when summated, equal ΣF_o. A second restriction must obviously be placed. It may be written

$$\sum F_o X = \sum F_t X, \tag{22}$$

or

$$\sum X(F_o - F_t) = 0, \tag{22a}$$

where X is the measure associated with the cell. Stated in more familiar verbal terms, the restriction is that the means of the observed and hypothetical arrays of measures shall be equal.

There is still too much freedom for values of F_t, if a normal curve is being fitted. The "scatter" of the hypothetical measures must also be restricted. In symbols,

$$\sum F_o X^2 = \sum F_t X^2, \tag{23}$$

or

$$\sum X^2(F_o - F_t) = 0. \tag{23a}$$

These equations, in effect, state that the variances of the observed and hypothetical arrays shall be the same.

The three restrictions represented by equations [21], [22], and [23] are imposed when a normal curve is fitted to a set of observed frequencies.[33] The equations are all linear *and* homogeneous in $(F_o - F_t)$. This is best seen from a comparison of equations [21a], [22a], and [23a].

Examples of non-linear restrictions. As a general rule, the restrictions imposed in applications of the chi-square test meet the linearity requirement. However, there are a few situations where non-linear restrictions are made. One of these has already been mentioned in the discussion of the phi-gamma function. In the phi-gamma case, theoretical frequencies are commonly obtained through a process that involves the minimizing of the sum of the squares of the differences between empirical and theoretical values of γ, where γ and p are non-linearly related. This non-linear relationship is clearly indicated by equation [15]. Despite the lack of linearity, the chi-square test is often applied. It must be realized that this test is not a rigid one; any estimate of χ^2 obtained from the differences between observed and theoretical values of p is not necessarily distributed as χ^2. Consequently, whenever χ^2 is used in evaluating the goodness of fit of the phi-gamma function, conclusions regarding the fit must be made with caution.

---

[33] One exception to this generalization should be mentioned. If values of the hypothetical mean and hypothetical variance are not estimated from the empirical data but come from some extraneous source, the only restriction is that $\sum F_o = \sum F_t$.

Another situation where a non-linear restriction is imposed in calculating theoretical proportions is in the use of Thurstone's Case V in treating data obtained by the method of paired comparisons (**13**, Chap. VII). The observed proportions are based on comparative judgments, the various stimuli being compared with each other. The proportions are translated into normal deviates (that is, into z-scores). Several z-scores typically enter into the calculation of each scale separation (distance). A scale separation based on several z-scores may be reconverted into a single "theoretical" z-score, the z-score then being used to obtain a theoretical proportion. A chi-square value computed from the differences between observed and theoretical proportions would not necessarily be distributed as χ^2. Ordinarily, of course, the chi-square test is not applicable to paired comparisons data because of a lack of independence between the observed proportions.

Non-linear restrictions are always imposed in obtaining theoretical frequencies or proportions if the reduction (transformation) process is used in curve fitting. (Almost any method of estimating parameters for complex functions may involve non-linear restrictions.) To illustrate: Grant and Norris (**10**) fitted a logarithmic function to the percentages in Table XXIV. In doing this, they probably transformed values of the independent variable (that is, values of It) into their logarithmic equivalents and used these in estimating the parameters for the logarithmic function. If they followed such a procedure, the calculated theoretical percentages were dependent upon a non-linear restriction.

Unfortunately, there is no way of determining the exact influence of non-linear restrictions on estimated values of χ^2. Therefore, if the chi-square test is applied, despite the imposition of non-linear restrictions, the investigator must be extremely cautious in interpreting the results,[34] bearing in mind that, as Cramér shows (**5**, Ch. 30), the calculated value of χ^2 is probably somewhat larger than it would be if the restrictions were linear.

CONCLUSIONS

Most readers will by now have correctly concluded that the chi-square test has a restricted usefulness. However, it usually cannot be replaced in those situations where it is applicable and it thus stands as a valuable research tool. Perhaps the chief trouble is that the test is too often applied without adequate prior planning; it is frequently "hit upon" and adopted after data have been collected and sometimes after other techniques of statistical analysis have been found unproductive.

---

[34] These statements hold also when the F statistic, based as it is on two independent estimates of χ^2, is applied in testing goodness of fit. For example, several of the F-tests proposed by Lindquist (**22**) depend for their exactness on the type of restrictions placed in estimating the parameters of the fitted functions.

The aim of every investigator should be to plan, in advance, not only every detail of every experiment but every step in the analysis of the anticipated data. All contingencies cannot be foreseen; but if the chi-square test is to be employed, there is no good reason for failing to provide for independence among the measures and for frequencies of adequate size.

There should seldom if ever be any compromising on the requirement of independence.[35] There should usually be no compromising on the size of frequencies. There are occasions, of course, when it is very time-consuming and perhaps very expensive to add more cases to a mere handful. The best procedure under such circumstances is to try for an experimental design which utilizes each subject to the limit and leads to an analysis of data on an individual rather than a group basis. If it turns out that only a few subjects can be studied and the data on each one cannot be analyzed separately, it may still be possible to find a method of analysis which is more exact than the chi-square test. For example, if the data can be arranged into a 2×2 table and the individual cell frequencies are less than 10, the *exact treatment* proposed by Fisher (7, pp. 96–97) is to be preferred to the chi-square test. The treatment is rather tedious to apply, but in view of its exactness there is no adequate excuse for avoiding it.

Many users and would-be users of the chi-square test gain erroneous impressions from what they read about limitations on the size of theoretical frequencies. A textbook says that frequencies of less than 10 are to be avoided. This statement is often interpreted to mean not that 10 is a limiting value to be exceeded whenever possible, but that 10 is a value around which the various theoretical frequencies may fall; and if an occasional frequency happens to be as low as 4 or 5, that is all right because other frequencies will be larger than 10 and everything will average out in the end. A textbook that gives 5 as the suggested minimum tends to encourage the retention of impossibly small theoretical frequencies. And so does a text which states, in effect, that Yates' correction for continuity should be applied if the cell frequencies are 5 or less and precision is desired. This implies not only that frequencies of less than 5 are quite acceptable, but also that Yates' correction is an antidote for small frequencies. Both implications are fallacious.

The following excerpts from Yule and Kendall (35, p. 422) may help to dispel false notions concerning the size of theoretical frequencies and also concerning the size of N:

---

[35] Any investigator who applies the chi-square test to interdependent frequency data should always feel obligated to include, in published accounts of his findings, a full explanation of the procedures employed and a justification of them.

> In the first place, N must be reasonably large. . . . It is difficult to say exactly what constitutes largeness, but as an arbitrary figure we may say that N should be at least 50, however few the number of cells.
>
> No theoretical cell frequency should be small. Here again it is hard to say what constitutes smallness, but 5 should be regarded as the very minimum, and 10 is better.

Hoel (**14**, p. 191), while giving 5 instead of 10 as the recommended minimal value of F_t, nevertheless emphasizes the importance of having a fairly large value of N by stating that if the number of cells or categories is less than 5, the individual theoretical frequencies should be larger than 5. Cramér (**5**, p. 420 f.) firmly recommends a minimal value of 10 and says that if the number of observations is so few that the theoretical frequencies, even after grouping, are not greater than 10, the chi-square test should not be applied. In all but one of the illustrations used by Cramér, the theoretical frequencies are considerably larger than 10, and in the exceptional case, he admits (p. 440) that the frequencies are smaller "than is usually advisable." An investigator handicaps himself whenever he applies the chi-square test in relation to small theoretical frequencies.

There are a few applications of the χ^2 test which have not been described and illustrated in the present survey, either because they are quite specialized in character or because they provide only approximate solutions. One of the specialized applications which may be of interest to some readers is Bartlett's test of the homogeneity of variance (**3**). Those who do not have access to Bartlett's original discussion will find a description of the test in Snedecor (**31**, pp. 249–251). Another specialized application of interest is the use of χ^2 in setting the confidence limits for a population variance from a known sample variance. The procedure is explained by Hoel (**14**, pp. 138–140) and need not be included here.

In general, any suggested applications of χ^2 which deviate from the well-established tests should be avoided except by those qualified to evaluate their full import or upon the advice of an expert.

BIBLIOGRAPHY

1. Anastasi, Anne, & Foley, J. P., Jr. An experimental study of the drawing behavior of adult psychotics in comparison with that of a normal control group. *J. exp. Psychol.*, 1944, **34**, 169–194.
2. Arnold, Magda B. Emotional factors in experimental neuroses. *J. exp. Psychol.*, 1944, **34**, 257–281.
3. Bartlett, M. S. Properties of sufficiency and statistical tests. *Proceed. Royal Soc.*, 1937, *A* **160**, 268–282.
4. Chen, H. P., & Irwin, O. C. Development of speech during infancy: Curve of differential percentage indices. *J. exp. Psychol.*, 1946, **36**, 522–525.
5. Cramér, H. *Mathematical methods of statistics.* Princeton: Princeton Univ. Press, 1946.
6. Edwards, A. L. *Statistical analysis for students in psychology and education.* New York: Rinehart, 1946.
7. Fisher, R. A. *Statistical methods for research workers* (10th Ed.). Edinburgh: Oliver and Boyd, 1946.
8. Fry, T. C. The chi-square test of significance. *J. Amer. stat. Ass.*, 1938, **33**, 513–525.
9. Goodfellow, L. D. A psychological interpretation of the results of the Zenith Radio experiments in telepathy. *J. exp. Psychol.*, 1938, **23**, 601–632.
10. Grant, D. A., & Norris, Eugenia B. Dark adaptation as a factor in the sensitization of the beta response of the eyelid to light. *J. exp. Psychol.*, 1946, **36**, 390–397.
11. Greenhood, E. R., Jr. *A detailed proof of the chi-square test of goodness of fit.* Cambridge: Harvard Univ. Press, 1940.
12. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942.
13. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
14. Hoel, P. G. *Introduction to mathematical statistics.* New York: Wiley, 1947.
15. Holt, R. B. Level of aspiration: Ambition or defense? *J. exp. Psychol.*, 1946, **36**, 398–416.
16. Irwin, O. C., & Chen, H. P. Development of speech during infancy: Curve of phonemic types. *J. exp. Psychol.*, 1946, **36**, 431–436.
17. King, H. E., Landis, C., & Zubin, J. Visual subliminal perception where a figure is obscured by the illumination of the ground. *J. exp. Psychol.*, 1944, **34**, 60–69.
18. Kuenne, Margaret R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
19. Lewis, D. *Quantitative methods in psychology.* Ann Arbor: Edwards Bros., 1948.
20. Lewis, Helen B. An experimental study of the role of the ego in work. I. The role of the ego in cooperative work. *J. exp. Psychol.*, 1944, **34**, 113–126.
21. Lewis, Helen B., & Franklin, Muriel. An experimental study of the role of the ego in work. II. The significance of task-orientation in work. *J. exp. Psychol.*, 1944, **34**, 195–215.
22. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
23. Lindquist, E. F. *Summary report of results.* Iowa City: Sixth Annual Every-Pupil Testing Program, 1934.
24. Pearson, K. Experimental discussion of the (χ^2, p) test for goodness of fit. *Biometrika*, 1932, **24**, 351–381.
25. Peatman, J. G. *Descriptive and sampling statistics.* New York: Harpers, 1947.
26. Peters, C. C., & Van Voorhis, W. R.

Statistical procedures and their mathematical bases. New York: McGraw-Hill, 1940.

27. Pronko, N. H. An exploratory investigation of language by means of oscillographic and reaction time techniques. *J. exp. Psychol.*, 1945, **35**, 433–458.
28. Seward, J. P., Dill, Jane B., & Holland, Mildred A. Guthrie's theory of learning: A second experiment. *J. exp. Psychol.*, 1944, **34**, 227–238.
29. Shaw, F. J., & Spooner, Alice. Selective forgetting when the subject is not "ego-involved." *J. exp. Psychol.*, 1945, **35**, 242–247.
30. Smith, S. The essential stimuli in stereoscopic depth perception. *J. exp. Psychol.*, 1946, **36**, 518–521.
31. Snedecor, G. W. *Statistical methods* (4th Ed.). Ames, Ia.: Iowa State College Press, 1946.
32. Stevens, S. S., Morgan, C. T., & Volkmann, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.*, 1941, **54**, 315–335.
33. Voelker, C. H. Phonetic distribution in formal American pronunciation. *J. acous. Soc. Amer.*, 1934, **5**, 242–246.
34. Yates, F. Contingency tables involving small numbers and the χ^2 test. *Suppl. J. Royal stat. Soc.*, 1934, **1**, 217–235.
35. Yule, G. U., & Kendall, M. G. *An introduction to the theory of statistics* (12th Ed.). London: Griffin, 1940.

Received June 19, 1949.

TASTE BLINDNESS TO PHENYL-THIO-CARBAMIDE AND RELATED COMPOUNDS

JOZEF COHEN AND DONALD P. OGDON
University of Illinois

I. THE PROBLEM

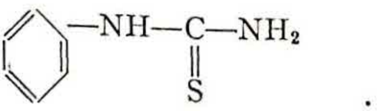
In 1931, Dr. A. L. Fox of the E. I. du Pont de Nemours Company was preparing a quantity of phenyl-thio-carbamide (hereafter referred to as P.T.C.) when a gust of wind blew some of the compound about. Dr. C. R. Noller, working in the same laboratory, complained of the bitter taste. Fox believed Noller to be mistaken, and put some of the crystals on his tongue. The taste sensation was neutral. Noller tasted some of the crystals and emphatically insisted that the compound was extremely bitter.

This incident touched off a series of experiments concerning the relationship of this and similar compounds to individual gustatory discrimination. Persons failing to taste the substance were termed "taste blind," following an analogous anomaly in color vision.

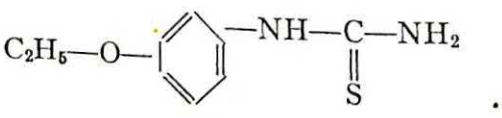
Fox (**43**, **44**) reported that taste blind observers constitute about 40% of the Caucasian American population. This figure was confirmed in approximation by Blakeslee and Salmon (**7**), Snyder (**37**, **38**), Fernberger (**15**), Blakeslee and Fox (**6**), and Parr (**28**). (See Table I.)

As it happened, the fact that some substances may taste differently to different people was not discovered by Fox, but by Laselle and Williams (**22**, **42**) in 1926. These last investigators noted that creatine, a muscle constituent, was as tasteless as chalk to some people, and very bitter to others. They noted that one pound of lean meat may contain close to two grams of creatine, and it is probable that this would cause meats to have a distinctive flavor to the taster. This taste difference would be especially pronounced in soups made from lean meats which must contain considerable amounts of creatine.

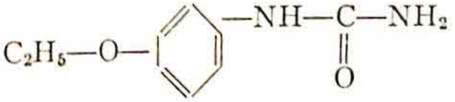
The structural formula of P.T.C., sometimes called phenyl-thio-urea, is given by



Para-ethoxy-phenyl-thio-carbamide is a substance closely related to P.T.C., and bears the same taste characteristics. Its structural formula is



In the above, if the single sulphur atom is replaced by an atom of oxygen, a commercial sweetening agent called "Dulcin" is obtained, whose structural formula is



.

Dulcin is used interchangeably with sugar in the manufacture of candy, with the difference that 250 times as much sugar as Dulcin is needed to achieve an equivalent sweetening effect. The similarity in chemical structure between P.T.C. and Dulcin was reported by Fox on several occasions (**43**, **44**, **16**, **17**).

Indeed, it was this close relationship which encouraged Fox (**17**) to investigate the entire family of compounds to determine the chemical bases of taste. Elaborate experimental evidence shows that it is the CS linkage which determines the taste of the substance. Fox (**16**) reports taste differences for many related mono-aryl-thio-carbamides and symmetrical di-aryl-carbamides. Of these compounds, only two have been found which deviate from the class to which they belong, but Fox attributes this deviation to insolubility. Fox (**43**) and Riddell and Wybar (**29**) report similar taste differences for the entire thiourea group, and extensive work with di-phenyl-guanidine is reported by Snyder and Davidson (**40**). It is interesting to note that tetra-methyl-thio-carbamide is reported to have a sour taste to both tasters and non-tasters, while thio-carbamide itself is reputedly nauseating, neither bitter nor sour to both tasters and non-tasters.

Blakeslee and Fox (**6**) and Blakeslee (**5**) report that there are a certain few individuals for whom P.T.C. tastes sweet. Almost all investigators report a small portion of the population for whom P.T.C. tastes sour, salty, camphory, sulphury, etc. That these and similar reports are an error of experimental procedure was suggested by Blakeslee (**14**) and Blakeslee and Fox (**6**) who insisted that these could more accurately be described or classified as odors.

II. PHYSIOLOGICAL FACTORS

Threshold Data. Blakeslee and Salmon (**8**) have reported that P.T.C. has the widest range of thresholds of any substance studied. It has been clearly established (for several racial groups) that the distribution of tasters' thresholds forms a bi-modal distribution. This evidence is reported by Blakeslee and Salmon (**7**), Blakeslee (**3**), Blakeslee and Fox (**6**), Parr (**28**), Rikimaru (**32**, **33**, **34**), Setterfield, Schott and Snyder

(36), and Matson (26). As an illustration, we have plotted the data of Setterfield, Schott, and Snyder (36) in Fig. 1.

The individual threshold variability to P.T.C. is very large. Blakeslee and Salmon (8) disclose experiments with ten subjects who were examined four times a day during from four to nine days. Thresholds for a given individual were found to vary by a factor of from 2 to 256. Further experiments testing seven individuals at fifteen-minute intervals indicated that changes occur in this short time—and even in a time as short as five minutes. No regularity was found in either the direction or the amount of change, nor was there any rhythm in the changes.

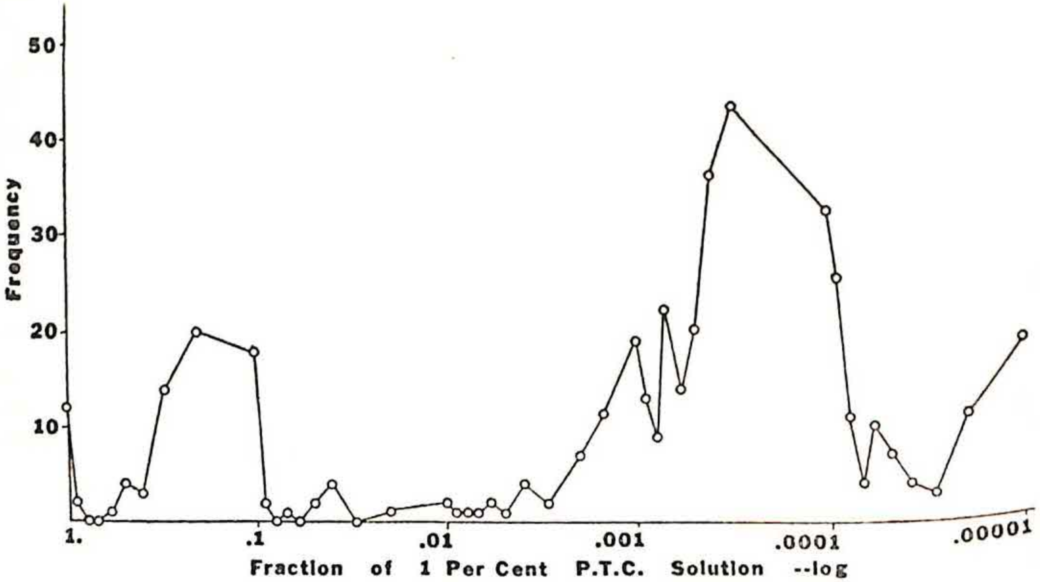


FIG. 1. Bimodal distribution of taste thresholds among tasters to P.T.C. The population is Caucasian. The number of individuals with a threshold of less than .00001 of 1% P.T.C. solution is 13. These data are plotted from tables given by Setterfield, Schott and Snyder (36).

Blakeslee (3) and Cook (14) note that threshold measurements for P.T.C. do not correlate with threshold measurements for quinine, picric acid, hydrochloric acid, saccharine, or salt.

Temperature in itself has no effect on the taste characteristics of P.T.C. When the solution is cold, however, less P.T.C. can be dissolved in it, and Blakeslee and Salmon (8) report a delayed reaction to such solutions. Analogously, Blakeslee (3) reports that many non-tasters may taste a hot solution, since more P.T.C. may be dissolved at high temperatures.

Saliva Condition. The saliva condition of the tasters and non-tasters has been the subject of much experimentation. Blakeslee (2), Fox (17), Mee (27), Blakeslee and Salmon (8) have all suggested that it may be the saliva which is the determiner of the taste or non-taste and not the "taste apparatus" itself. Fox (17) specifically suggested the possibility that non-tasters have in their saliva a product, possibly a protein or a colloid, which precipitates this compound as a very insoluble substance which causes no taste sensation.

Snyder (37, 38) and Blakeslee (3) report that there is no correlation between the rhythmic change in the pH of the saliva and taste acuity. Blakeslee and Salmon (8) dried the tongues of each of their subjects with blotting paper and then rinsed these tongues with water. Four of six subjects were rendered less sensitive after this treatment, and the degree of desensitization varied considerably among the four testees.

The problem of the influence of saliva was analyzed by Cohen and Ogdon (13) in 1949. They reasoned that this question might be settled by allowing tasters and non-tasters to taste using *another* taster's or non-taster's saliva. Their results indicate that an individual will have the taste sensation when the following two necessary conditions are met: (1) he must have the correct "taste apparatus" and (2) he must have his own saliva. A non-taster cannot taste under any conditions, even when he uses the saliva of a taster. A taster cannot taste under any circumstances, except when he uses his *own* saliva; he cannot taste if he uses the saliva of another taster or non-taster. He can taste if he uses his own saliva, even though the saliva is placed on his tongue in exactly the same manner as the saliva from another individual. No observer can taste P.T.C. when the crystals are dissolved in water and no saliva is used at all.

By way of theoretical explanation, salivas are probably as different as fingerprints. Blakeslee and Salmon indicate that the saliva is important and Cohen and Ogdon note the fact that saliva is necessary to taste P.T.C. The observation that it must be the individual's own saliva may be brought about by the fact that the "taste apparatus" becomes, over the years, extremely sensitive and specialized to the particular saliva which the individual possesses. This being the case, when other saliva is introduced, it is equivalent to water, and no taste sensation results. Additional experiments by Cohen and Ogdon indicate that saliva *may* aid in the taste of other substances, but this effect is most pronounced with P.T.C.

Sex. The sex factor in the ability to taste P.T.C. has been investigated with conflicting results. Originally, Fox (43, 44) reported that taste deficiency was not related to sex, and this was confirmed by Snyder

(37, 38), Rikimaru (32, 33, 34), Setterfield, Schott and Snyder (36), and Snyder and Davidson (40). On the other hand, Fernberger (15) and Boyd and Boyd (11, 12) report the taste ability between the sexes to be significant. These last investigators published the ratio of female tasters to male tasters in Wales to be 1.66. The Fernberger ratio is only 1.11. Slight sex differences are noted in most other papers, but they are statistically insignificant.

Age. The possibility of age influencing the ability to taste P.T.C. has been investigated by Blakeslee and Salmon (7), Rikimaru (32, 33, 34), and Snyder (37, 38). On the whole, the age factor may be considered to be unrelated to taste ability. The apparent slight increase in acuity at certain ages, reported by Rikimaru, is probably insignificant.

Smoking. Hall and Blakeslee (20) report extensive experiments concerning the effect of smoking on ability to taste P.T.C. These experimenters set down the following conclusions: (1) Smoking has a definite effect upon taste acuity. (2) The time which must elapse after smoking before the threshold returns to a resting position varies with the individual. In these experiments only 58% returned to normal within one hour. (3) The incipient effect of smoking is the direct dulling of the "taste apparatus." (4) The true effect of nicotine on the nerves associated with taste appears to be the same as for other nerves—initial stimulation followed by depression.

III. Genetic and Environmental Factors

The ability to taste P.T.C. is inherited as a simple unit-factor dominant in accordance with the primary Mendelian principle. This is an extremely important characteristic of P.T.C. since it affords the geneticist a simple test of the Mendelian laws.

Family Studies. In 1931 Blakeslee and Salmon (8) and Snyder (35) independently discovered the simple recessive character of taste blindness. They classified the parents of families into three groups according to their phenotypes—both parents tasters, one parent taster, neither parent taster. The taster vs. non-taster characteristics of the offspring were measured, and the Mendelian nature of the transmission inferred.

In the following year, Snyder (38) conducted a monumental study wherein he took data concerning 800 families at the Genetics Laboratory of the Ohio State University. His theory concerning the simple Mendelian inheritance was confirmed. Other confirmation using families has come from Rikimaru (32, 33, 34, 35), Strandskov (41) and Levit and Soboleva (25).

Twin Studies. Twenty pairs of twins, classified as identical by relatives, were studied by Rife (30). Each pair of twins had the same taste characteristics with the exception of one pair of apparently monozygotic twins. But in this case the non-taster twin had suffered a general taste deficiency, which may have been the result of a severe case of poliomyelitis. Ardashnikov, Lichtenstein, Martynova, Soboleva and Postnikova (**1**) investigated 137 pairs of identical twins and found only three pairs who differed in their reactions to P.T.C. The non-taster sibling of one of these pairs was pathologically handicapped. However, no physiological or psychological anomaly was reported for the other two pairs who differed in reaction.

A somewhat more elaborate twin study has been conducted by Botsztejn (**10**) wherein he divided his population into three groups—tasters, taste weak, and taste blind. Using groups of identical and fraternal twins, he concluded that hereditary transmission followed a multiple allele series. Taster was dominant to taste weak, and taste weak was dominant to taste blind.

Transmission of Taste to Related Compounds. Snyder and Davidson (**40**) report that the inheritance of taste to di-phenyl-guanidine follows the same laws as P.T.C. Riddell and Wybar (**29**) conclude that taste to thiouracil is inherited in the same manner.

Linkage in Transmission. All investigators report that taste ability to P.T.C. is neither sex-linked nor sex influenced. Snyder, Baxter, and Knisley (**39**) have found that reaction to P.T.C., blood groups, and blood types are all inherited independently, nor is it related to brachydactyly (**21**).

Environmental Factors. Blakeslee (**3**) has suggested that taste to P.T.C. may be modified by environmental factors, but no evidence has been secured to confirm this.

IV. ETHNOLOGICAL FACTORS

The simple inheritance characteristics of taste to P.T.C. have encouraged anthropologists to attempt the tracing of ethnological groups by the per cent taster in various populations. It was hoped that the numbers would form rough racial clusters in the same manner as blood groups.

A summary of the extensive investigations by interested scientists is given in Table **1**. Unfortuntely, no ordering seems to result. Indeed the data themselves are internally inconsistent. Witness the range of American Caucasians from 60% to 82%. This indicates that the experimental procedures are weak, and some standard method of administration of P.T.C. should be agreed upon. Almost all of the studies in Table 1 merely describe populations and give results; there is little discussion of experimental technique.

TABLE I

PER CENT OF TASTERS TO P. T. C. IN VARIOUS POPULATIONS

Group	*Per Cent Taster*	*N*	*Investigator*
Caucasians (American)	60		Fox (**43, 44**)
Arabs (Syria)	63	400	Hudson and Peter (**28**)
Armenians (Syria)	68	294	Berberian (**28**)
Northern Jews (Palestine)	68	245	Yunovitch (**28**)
Semenites (Palestine)	68	59	Yunovitch (**28**)
Caucasians (American)	68	283	Blakesleee and Salmon (**7**)
Caucasians (American)	69	439	Parr (**28**)
Caucasians (American)	69	440	Snyder (**37**)
Caucasians (American)	70	3,643	Snyder (**38**)
Southern Jews (Palestine)	72	175	Yunovitch (**28**)
Caucasians (American)	73	1,025	Fernberger (**15**)
Egyptians	76	208	Hickman & Marcos (**28**)
Negroes (Alabama)	77	533	Howard & Campbell (**28**)
Caucasians (American)	80	232	Rife & Schonfeld (**31**)
Caucasians (American)	82	477	Setterfield, Schott, & Snyder (**36**)
Caucasians (Jews)	85	82	Rife & Schonfeld (**31**)
Formosans (Chinese origin)	89	5,933	Rikimaru (**32, 33, 34**)
U. S. Indians (mixed blood)	90	110	Levine and Anderson (**24**)
Negroes (American)	90	3,156	Lee (**23**)
Negroes (Kenya)	91	110	Lee (**23**)
Japanese	91	921	Fukoka (**18**)
Negroes (American)	92	107	Setterfield, Schott, & Snyder (**36**)
Japanese	93	8,824	Rikimaru (**32, 33, 34**)
Chinese	94	167	Chen and Chain (**28**)
U. S. Indians (full blood)	94	183	Levine & Anderson (**24**)
Formosans (Aborigines)	95	1,756	Rikimaru (**32, 33, 34**)
Egyptian Sudan natives	96	805	Lee (**23**)
Koreans	97	55	Fukuoka (**18**)

V. SUMMARY

This paper has discussed the taste blindness of certain individuals to phenyl-thio-carbamide and related compounds. Physiological, genetic, environmental, and ethnological factors have been considered.

BIBLIOGRAPHY

1. ARDASHNIKOV, S. N., LICHTENSTEIN, E. A., MARTYNOVA, R. P., SOBOLEVA, G. V., & POSTNIKOVA, E. N. The diagnosis of zygosity in twins. *J. Hered.*, 1936, **27**, 465–468.

2. BLAKESLEE, A. F. Genetics of sensory thresholds: taste for phenylthiocarbamide. *Science*, 1931, **74**, 607.

3. BLAKESLEE, A. F. Genetics of sensory thresholds: taste for phenylthiocarbamide. *Proc. nat. Acad. Sci., Wash.*, 1932, **18**, 120–130.

4. Blakeslee, A. F. Some differences between people in taste and smell reactions. *Eugen. News*, 1933, **18**, 63.
5. Blakeslee, A. F. A dinner demonstration of threshold differences in taste and smell. *Science*, 1935, **81**, 504–507.
6. Blakeslee, A. F., & Fox, A. L. Our different taste worlds. *J. Hered.*, 1932, **23**, 97–110.
7. Blakeslee, A. F., & Salmon, M. R. Odour and taste blindness. *Eugen. News*, 1931, **16**, 105–109.
8. Blakeslee, A. F., & Salmon, T. N. Genetics of sensory thresholds: variation within single individuals in taste sensitivity for PTC. *Proc. nat. Acad. Sci., Wash.*, 1935, **21**, 78–83.
9. Blakeslee, A. F., & Salmon, T. N. Genetics of sensory thresholds: Individual taste reactions for different substances. *Proc. nat. Acad. Sci., Wash.*, 1935, **21**, 84–90.
10. Botsztejn, C. Zur Kenntnis der Geschmacksblindheit gegenüber Phenylthiocarbamid (P.T.C.) in der Zurcher Bevölkerung und deren Erbgang. *Archiv der Julius Klaus-Stiftung*, 1942, **17**, 109–123.
11. Boyd, W. C., & Boyd, L. G. New racial blood group studies in Europe and Egypt. *Science*, 1936, **84**, 328–329.
12. Boyd, W. C., & Boyd, L. G. Sexual and racial variations in ability to taste phenyl-thio-carbamide, with some data on the inheritance. *Ann. Eugenics*, 1937, **8**, 46–51.
13. Cohen, J., & Ogdon, D. P. Taste blindness to phenyl-thio-carbamide as a function of saliva. *Science*. (In press.)
14. Cook, R. C. Inherited variations in the sense of taste. *American Mercury*, 1933, **28**, 67–69.
15. Fernberger, S. W. A preliminary study of taste deficiency. *Amer. J. Psychol.*, 1932, **44**, 322–326.
16. Fox, A. L. The relationship between chemical composition and taste. *Science*, 1931, **74**, 607.
17. Fox, A. L. The relationship between chemical constitution and taste. *Proc. nat. Acad. Sci., Wash.*, 1932, **18**, 115–120.
18. Fukuoka, G. Frequency of taste-blindness among the Japanese and related races. *Eugen. News*, 1936, **21**, 52–54.
19. Gottschick, J. Erbliche Unterschiede der Geschmackemfindungen auf P-Aethoxyphenylthioharnstoff. *Zeitschrift für Menschliche Vererbungs- und Konstitutionslehre*, 1937, **21**, 254–265.
20. Hall, A. R., & Blakeslee, A. F. Effect of smoking on taste thresholds for phenyl-thio-carbamide (P.T.C.). *Proc. nat. Acad. Sci., Wash.*, 1945, **31**, 390–396.
21. Hogben, L., & Pollack, R. A contribution to the relation of the gene loci involved in the isoagglutinin reaction, taste blindness, Friedreich's ataxia and major brachydactyly of man. *J. Genetics*, 1935, **31**, 353–361.
22. Lasselle, P. A., & Williams, R. J. The identification of creatine. *J. Amer. chem. Soc.*, 1926, **48**, 536–537.
23. Lee, B. F. A genetic analysis of taste deficiency in the American Negro. *Ohio J. Sci.*, 1934, **34**, 337–342.
24. Levine, P., & Anderson, A. S. Observation on taste blindness. *Science*, 1932, **75**, 497–498.
25. Levit, S. G., & Soboleva, G. V. Comparative intrapair correlations of fraternal twins and siblings. *J. Genetics*, 1935, **30**, 389–396.
26. Matson, G. A. Blood groups and ageusia in Indians of Montana and Alberta. *Amer. J. phys. Anthrop.*, 1938, **24**, 81–89.
27. Mee, A. J. Taste and chemical constitution. *Sci. Progr. Twent. Cent.*, 1934, **29**, 228–235.

28. Parr, L. W. Taste blindness and race. *J. Hered.*, 1934, **25**, 186–190.
29. Riddell, W. J., & Wybar, K. C. Taste of thiouracil and phenylthiocarbamide. *Nature, Lond.*, 1944, **154**, 669.
30. Rife, D. C. Genetic studies of monozygotic twins. *J. Hered.*, 1933, **24**, 339–345.
31. Rife, D. C., & Schonfeld, M. D. A comparison of the frequencies of certain genetic traits among Gentile and Jewish students. *Hum. Biol.*, 1944, **16**, 172–180.
32. Rikimaru, J. [A study of "taste blindness." The incidences among Japanese and Formosans, and divergencies in race, age, and sex.] *Jap. J. Psychol.*, 1934, **9**, 87–108.
33. Rikimaru, J. [A study of the individual and racial differences in taste reaction using the same stimulus compound.] *Jap. J. Psychol.*, 1934, **9**, 901–921.
34. Rikimaru, J. Taste deficiency of Japanese and other races in Formosa. *Amer. J. Psychol.*, 1936, **48**, 649–653.
35. Rikimaru, J. [Taste differency for P.T.C., with special reference to its hereditary nature.] *Jap. J. Psychol.*, 1937, **12**, 33–54.
36. Setterfield, W., Schott, R. G., & Snyder, L. H. Studies in human inheritance. XV. The bimodality of the threshold curve for the taste of phenyl-thio-carbamide. *Ohio J. Sci.*, 1936, **36**, 231–235.
37. Snyder, L. H. Inherited taste deficiency. *Science*, 1931, **74**, 151–152.
38. Snyder, L. H. Studies in human inheritance. IX. The inheritance of taste deficiency in man. *Ohio J. Sci.*, 1932, **32**, 436–440.
39. Snyder, L. H., Baxter, R. C., & Knisley, A. W. Studies in human inheritance. XIX. The linkage relations between blood groups, the blood types, and taste deficiency to P.T.C. *J. Hered.*, 1941, **32**, 22–25.
40. Snyder, L. H., & Davidson, D. F. Studies in human inheritance. XIII. The inheritance of taste deficiency to di-phenyl-guanidine. *Eugen. News.*, 1937, **22**, 1–2.
41. Strandskov, H. H. The distribution of human genes. *Sci. Mon., N. Y.*, 1941, **52**, 203–215.
42. Williams, R. J. "Taste deficiency" for creatine. *Science*, 1931, **74**, 597–598.
43. *Anon. Science*, 1931, **73**, suppl. 1894, 14.
44. *Anon.* Six in ten "tasteblind" to bitter chemical. *Sci. News Lett.*, 1931, **19** 249.

Received May 27, 1949.

ARMY AIR FORCES AVIATION PSYCHOLOGY PROGRAM RESEARCH REPORTS[1]

SPECIAL REVIEWS

FLANAGAN, JOHN C. *The aviation psychology research program in the Army Air Forces.* AAF Aviation Psychology Program Research Reports, Report No. 1. Washington, D. C.: U. S. Government Printing Office, 1948. Pp. xii+316. $2.00.

On July 15, 1941, John C. Flanagan was commissioned as a major in the Army's Officers' Reserve Corps and was at once assigned to duty in the Medical Division, Office, Chief of Air Corps. This was the beginning of the extensive research enterprise which was to become the Aviation Psychology Program of the Army Air Forces. Four years later the program included some 200 officers, 750 enlisted men, and 500 civilians. At that time, Colonel Flanagan could look back with pardonable pride upon a distinguished record of leadership in research planning, organization, and application. The comprehensive account of this work in the nineteen volumes of research reports under review in this issue of the PSYCHOLOGICAL BULLETIN symbolizes the effective manner in which the program developed under the direction of Colonel Flanagan and his many able associates. In historical perspective, this collective research effort will stand out as an important landmark in the evolution of applied psychology, as well as a very substantial contribution towards the effectiveness of the Army Air Forces in World War II.

In the first volume of the series, Flanagan presents a general account of the entire program, in three main parts: (1) a record of the expanding scope of the program, in terms of problems studied and of official directives on organization and functions; (2) a summary of the principal research findings and their applications to AAF problems; (3) an appraisal of the general contributions to theory, knowledge, and techniques of aviation psychology. The second section is a condensation of the results reported in the other eighteen volumes, which are reviewed individually in subsequent pages of this issue.

The official account of the expansion of the program is most impressive. In the beginning, research was restricted to the selection of pilots. Soon its scope was broadened to include bombardiers and navigators.

---

[1] *Editor's Note.* The nineteen volumes in this series were issued over a period of more than two years, the last report appearing in the spring of 1949. In spite of the long delay in the completion of the series, due partly to printing difficulties, it was decided to postpone the publication of any of the reviews until all of them were available. The advantages of simultaneous and somewhat coordinated evaluation of the entire program seemed to overweigh the possible decline of interest in these notable war-time research achievements with the passage of time.

Eventually gunners were added, and so the selection and classification of all air-crew personnel became the responsibility of the psychological research units. Attention then turned to the complex field of training, in which some of the most outstanding work in the program was done. In late 1943 interest in operational problems was developing, since the ultimate validation of selection, classification, and training procedures could only be found in combat performance. In May 1944 an evaluation and research detachment was flown to the European theatre to study combat operations. Permanent research units for combat commands were later authorized, but the war ended before they could go overseas. Finally, at the end of the war, the conception of an integrated research organization had crystallized which would study selection and classification, training, utilization of personnel, and the human factor in the design and operation of equipment. Although this plan, adopted in 1946, was to be largely supplanted two years later by a more comprehensive organization, these earlier recommendations of AAF psychologists laid the foundations for an integrated program of research and development in the Air Force.

In the third part of the book, Flanagan summarizes the general scientific and technological achievements of the AAF psychological research program. Like all other wartime research efforts, the gains were far greater in the technological than in the scientific dimension. One looks in vain for really significant scientific progress in any of the areas discussed; immediate military necessity would not permit long-range, basic research.

The classification research made possible extensive studies of the interrelations of traits, which seemed to support an atomistic theory of psychological organization. The factor analyses and the low correlations among apparently quite similar activities both indicated that "the number of basic uncorrelated aptitude factors . . . (is) probably much larger than most experts had supposed." In gunnery training, for example, men who did well on one type of trainer might do very poorly on another which seemed quite similar. Because of such results, generalization and transfer of training find little place in Flanagan's psychological scheme of things. (Except in the emphasis upon correlations between test scores and a variety of fairly dissimilar criteria!)

In the field of training the most outstanding contribution was the determined effort to secure adequate criteria of proficiency against which both classification and educational procedures could be checked. The nature and the determinants of effective behavior on the job became the central concern of the program in its later stages. These activities were for the most part models of skillful and persistent attack upon these most difficult of all problems in applied psychology. Although definitive solutions to these problems were not as a rule forthcoming, the methodological gains were very substantial. Future workers in this field will know better how to approach such tasks, as well as what hypotheses and procedures to reject.

LYLE H. LANIER.

New York University.

DUBOIS, PHILLIP H. (ED.) *The classification program.* AAF Aviation Psychology Program Research Reports, Report No. 2. Washington, D. C.: Government Printing Office, 1947. Pp. xiv+394. $1.50.

DuBois' report is an important historical document. It is important only in part because much money was spent in the aviation psychology classification program—$1,309,821 is the estimated expenditure for 1943 alone (p. 305)—some of which may be recovered from this book and its supporting files in a future war emergency. It is important because some of the mistakes made in experimental design can, if viewed critically and constructively, point the way to more adequate research development of differential selection tools. (The book being reviewed does not, however, look critically upon the experimental designs employed in the studies.) Furthermore, the report is important for the implications it has for organization and implementation of needed personnel research programs of large scope in the area of industrial as well as military personnel.

There are six chapters and a brief summary chapter. The first chapter deals with planning for selection of trainees and differential classification of pilots, bombardiers and navigators. The second chapter discusses the AAF psychological organizations concerned with selection and classification of air crew. The organization of the various AAF research units and the ten medical and psychological examining units are treated in considerable detail.

In Ch. 3 the classification batteries are discussed. The first battery to be used in the examining units of the Air Surgeon was developed early in 1942. Five changes were made in it during that year, and four changes were made in the next three years. During most of the period the Air-Crew Classification Test Battery included approximately 14 objective-type printed tests and six apparatus tests.[1]

Of greatest interest to many personnel technicians is the manner of weighting the many tests comprising the classification battery:

> The classification battery of February, 1942, employed an arbitrary weighting system: " . . . the necessity for an immediate classification demanded an a priori system utilizing the tests that were available" (p. 76). With reference to the classification battery of June, 1942: "On the basis of further professional analyses of the requirements of air crew, the weighting system was considerably revised" (p. 86). For the classification battery of August, 1942: "For the first time, the classification battery was used for selection instead of *merely for*

---

[1] *Editor's Note.* These tests are described in the reviews of Reports 4 and 5.

classification. . . . The weights were readjusted on the basis of professional judgment, since validity data were still incomplete and the test intercorrelations were not available" (p. 90, italics the reviewer's). But in the classification battery of July, 1943: "for the first time test weights for all three air crew aptitude scores were close approximations to multiple regression weights based directly on empirical data" (p. 99). (This weighting system suggests evolutionary progress.) The classification battery for September, 1944, included predictor scores for a total of seven air crew specialties: bombardier, navigator, bomb pilot, aerial gunner, air mechanic-gunner and radio operator-gunner. A final war-time revision of the battery was made in June of 1945, to segregate B-29 gunners from other types of gunnery trainees. The reviewer believes the following quotation suggests that the 1945 battery has atavistic traits: "The weightings for this specialty were based upon job analyses and one test, the *Pedestal Sight Manipulation Test*, with high face validity for gunnery, was introduced into the battery, carrying 20 per cent of the weight for the specialty" (p. 103). The systems of weighting for *differential prediction* were, for the most part, quite arbitrary. It is here that a new look at experimental design can perhaps be of greatest value.

Chapter four presents results of validity studies concerned with relation of total weighted scores, called stanines, and non-test data such as age and education to training criteria.[2] The training criteria against which these variates were correlated were pass-fail criteria. The fail criteria were eliminations from training for flying deficiency, for fear, and elimination at trainee's own request. The weighted predictor scores when correlated against such pass-fail criteria for many classes comprising thousands of trainees yielded biserial correlations for most samples of pilots that were fairly high and nearly always significant; those for navigators were moderately high and frequently significant; those for bombardiers were low and frequently not significant. Even moderately low biserial correlations yield striking predictive efficiency when the results are set up in charts, as is done often in this chapter.

One of the crucial indices of effectiveness of a differential classification battery (as contrasted to one designed for *mere selection*) would seem to be validity of a battery weighted for one specialty for another specialty. For example, how well does the pilot battery work for bombardier, the navigator battery for pilot, the bombardier battery for navigator, etc. Judging from the results presented the answer appears to be that the battery has higher validity for the specialty for which it was designed. However the critical observer may ask for: (1) an experimental design specifically aimed at evidence for or against the efficacy of a battery in differential prediction as contrasted to selection; and (2) safeguards against artificial contamination of the criterion. Did the trainee know his stanine? Did the instructor know the trainee's stanine? If so, did such knowledge influence the criteria of elimination for flying deficiency, for fear, or for elimination at the trainee's own request?

---

[2] Validities of printed tests and of apparatus tests are not included in this book, they being given in Reports No. 4 and No. 5; validities of weighted tests when correlated with combat criteria are also presented in Report No. 17.

One table of information is presented in chapter four that is particularly damaging to the belief that stanines have considerable power in *differential prediction* as contrasted to mere selection. This information pertains to the validity of various stanines in predicting scores of trainees in pistol and carbine fire! For 350 cases the various air crew stanines were found to correlate from .25 to .40 when correlated with performance on the ground with the pistol, and from .21 to .31 with target performance with the carbine. The author refers to these results as evidence of "the versatility of the stanines" (p. 179).

Based on statistical and clinical approaches, Ch. 5 tells what happened to 1,311 air crew trainees who had been given tests but who were not screened by tests. For this experimental group the pilot stanine correlated .66 with the criterion. For the preliminary qualifying examination "only 45 out of 520 men who failed the examination were graduated from training, whereas the yield among those passing the examination was 211 out of 751" (p. 201). The study reported in this chapter is one of the most interesting to come out of wartime personnel research.

ROGER M. BELLOWS.

Wayne University.

THORNDIKE, ROBERT L. (ED.) *Research problems and techniques.* AAF Aviation Psychology Program Research Reports, Report No. 3. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. viii+163. $1.00.

A research program usually has two outcomes: (1) increased knowledge in the area being investigated and (2) increased understanding of the research methods used. It would have been a shame if the reporting of the large-scale AAF Aviation Psychology Program had been restricted to the first of these two outcomes. Fortunately, Report No. 3 was included in the series so that the psychological profession could share in the distillation of research experiences provided by that program.

Using the relatively unified aircrew selection and classification program as a frame of reference, Thorndike describes the practical research problems which arose and the adequacy of the methods of investigation applied. The report is organized around the major steps in a test validation program, separate chapters being devoted to job analysis, test development, criterion measurement, determination of test reliability and validity, and problems of combining and weighting tests when used in a battery. The influence upon test results of variables such as apparatus model, examiner, time of day, etc., is also discussed. After pointing out the problems common to both classification and training research, Thorndike adds a brief chapter on problems and techniques peculiar to the latter, such as administrative difficulties, equating of groups, control of training variables, etc.

If Thorndike's style of writing had a lower Flesch count one might classify this report as a "How To" manual. It is this, to the extent that it provides an excellent guide to the step-by-step procedures required in the validation of an aptitude testing program. The sources of information concerning job requirements are presented, methods for test development are described, types of criteria are compared, and statistical methods and computational routines are presented in some detail. As such, it will serve as a useful reference in personnel psychology.

To consider this report as merely a manual of procedures, however, would be a grave injustice. For it performs the much more valuable function of taking the reader "behind the scenes" in order to see the *Why* as well as the *How*. Procedures are evaluated as well as described, limitations as well as strengths are pointed out, and theoretical implications as well as practical considerations are presented. Some of the topics of particular interest to this reviewer were the following: (1) the value of job analysis versus trait analysis, (2) objective versus subjective criteria, (3) selection versus classification, (4) evaluation of various measures of reliability in terms of the sources of variance affecting each of the measures, (5) problems of restriction of range of ability due to prior selection. These and other basic problems are treated systematically and the rationale as well as the methods of attack are described.

The editor recognizes (and clearly states in his Preface) one limitation of the report, i.e., its restriction in large part to the selection and classification program. Another limitation, in the reviewer's opinion, was the failure to discuss in detail the study of a non-preselected "experimental group" given aircrew training regardless of their test scores. Although this project is described in Report No. 2 of this series, it warranted more than the brief reference in this report in connection with the problem of restriction of range. From a methodological point of view, it represented a major step in providing crucial evidence on the validity of the classification program, and a discussion of its rationale, procedures, and general implications for experimental design would have been appropriate.

ALBERT S. THOMPSON.

Teachers College, Columbia University.

MELTON, ARTHUR W. (ED.) *Apparatus tests.* AAF Aviation Psychology Program Research Reports, Report No. 4. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. xxi+1056. $3.50.

This report describes the largest program of psychomotor selection research yet undertaken. Although the National Research Council had conducted psychomotor selection projects on civilian flyers, little equipment or information was available and many psychologists believed that it was not feasible to administer individual apparatus tests to large numbers of candidates.

The tests developed emphasized measures of precision at high speed rather than steadiness, equilibrium, or strength. The standard classification battery at various times included: (1) complex coordination (Mashburn type); (2) two-hand coordination; (3) two-hand pursuit; (4) discrimination reaction time; (5) finger dexterity; (6) arm, hand steadiness and aiming; (7) pedestal sight manipulation. Complex pursuit or continuous coordinations appeared to be most valid for flying skills. Major advances in procedure included standardized instructions, automatic cycling of trials, and multiple unit controls for testing four persons under each experimenter. Further improvements in the design of tests and the correlation of findings on specific tests with specific reasons for success or failure in flying skills would seem to call for more intensive empirical and theoretical job analyses.

In addition to the development and validation of tests for the selection of pilots, bombardiers, navigators, gunners, radar operators, etc., most tests were later devised to emphasize some variation in the performance such as working under distraction, stress or various degrees of difficulty, but with little increase in validity. Early test trials, including practice trials, were often as valid or even better than additional trials at later stages of learning. Rates of learning on various tests were quite specific.

This suggests to the reviewer that initial work methods adopted by a candidate in a new situation were important despite standardized instructions. The greater validity of more complex tests suggests as a second explanation that ability to attend simultaneously to various aspects of a complex situation might have been more important than the motor aspects of performance. Tests in which distractions were usually quite separate from the main operation would not be an adequate test of this theory. The success of work sample tests involving coordinated movements of hands and feet in a central task would be more relevant. Unfortunately, the most representative work sample tests were not employed for classification purposes until relatively late when there was no longer an opportunity for large scale validation, but results were promising.

In a program which frequently "washed out" 35 per cent of those who had already passed medical examinations, intelligence tests, and several years of college work and in which the average cost of flying training was on the order of $25,000 per man, any reduction in "washouts" is important, particularly when psychomotor tests cost on the order of $5.00 per person. Bar graphs showing the percentage of persons failing the primary training from each range of scores show clearly the savings from the use of single psychomotor tests. Multiple correlations show that the battery of tests gave still greater validity. The saving of lives and equipment by elimination of trainees of lower aptitudes and the selection of abler trainees for actual combat duty are obvious advantages of a more strictly military nature.

To the reviewer this program has important implications as to needs in future research and training in psychology, e.g.: (1) The training of additional specialized personnel for *instrument design* and *construction*; (2) likewise, the relative neglect of *intensive job analysis* and *motion study* in psychomotor performances on military equipment calls for increased attention to this phase of training for future psychologists; (3) the further development of *more complete factorial batteries* of psychomotor tests would enable applied psychologists to select tests having the greatest probable validities for various skills; (4) the development of *procedures for setting up specific criteria* for success in acquiring each major type of skill; (5) the rapid mechanization of military aircraft and equipment, e.g., remote control aircraft and missiles, may outmode a great deal of present testing and require different aptitudes; (6) a *standby research and selection program, revised periodically*, is necessary to provide an adequate plan for emergencies. This in turn is likely to require an *administrative plan* which will encourage the widespread participation in research of civilian as well as military psychologists.

ROBERT H. SEASHORE.

Northwestern University.

GUILFORD, J. P. (ED.) *Printed classification tests.* AAF Aviation Psychology Program Research Reports, Report No. 5. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. xi+919.[1] $3.00.

Printed Classification Tests gives a detailed account of methods, results, and conclusions concerning paper and pencil tests used and considered for use in the selection of Army Air Force pilots, navigators, and bombardiers. The volume will be of interest not only to aviation psychologists but to all psychometricians because of the important factor analyses reported. It presents impressive evidence of the effectiveness of the organization, direction and continuity of purpose of the group of psychologists who worked in the AAF research program.

The methods followed in the development of the AAF tests were in the approved text-book style of test development that is almost never pursued so thoroughly in actuality: job analyses were made; hypotheses were clearly formulated for factors that would appear to be needed; tests were given before the groups were selected; factor analyses were made of tests and criteria; new tests were based on the validity and factor studies; the process was repeated. This is the first successful attempt to use factor analysis in a selection problem to show the factorial composition of the criterion as well as of the tests. The criterion unfortunately had to be only training school records rather than combat success. Factors extracted by the centroid method were rotated orthogonally to an approximate simple structure.

---

[1] This review has been read and constructive criticisms provided by Mr. George Burgess and Miss Dorothy Everett.

The tests tried were the commercial tests that looked most promising plus many new ones that were developed. Tests relating to these areas were examined: intellect and information; verbal ability; mathematics; reasoning; judgment, foresight and planning; integration; memory; visualization; mechanical; information; perception; size and distance estimation; spatial; orientation; set and attention; temperament and personality. In addition to tests the following kind of variables are correlated with training records: clinical type procedures, motivation, and biographical data.

> Twenty-seven separate factors, rarely the ones expected, were identified with varying degrees of clarity. Seventeen factors have positive validity for predicting success in pilot training pass-fail criterion. These seventeen factors are in order of their validity for the pilot criterion: (1) Spatial Relation, (2) Mechanical Experience, (3) Pilot Interest, (4) Integration I, (5) Visualization, (6) Psychomotor Coordination, (7) Perceptual Speed, (8) Length Estimation, (9) Memory III, (10) Judgment, (11) Mathematical Background, (12) Integration II, (13) Visual Memory, (14) Planning, (15) Paired Associates Memory, (16) Reasoning II, and (17) Spatial II.
>
> These factors range in estimated validity for the pilot training criterion from .32 for the Spatial Relations factor to .05. Since so many orthogonal factors are found in a single criterion it is to be expected that validities of the factors would be low. Furthermore the large number of cases used in these studies indicates that the validities reported are probably more dependable than some of the higher validities reported in industrial research studies.

In his chapter on "General Conclusion" the editor remarks that a complete classification battery should cover at least 25 factors. This probably overstates the case since the inclusion of those factors having lower validities does not contribute substantially to the validity of the composite. In addition, no more than 15 of the 27 factors are defined adequately enough for inclusion in a comprehensive classification battery.[2] The work that is reported is gigantic. There will be enough leads to explore for dozens of doctoral dissertations based on the factors found and suggested. Perhaps the most important new factor found is Judgment. Also of importance is that the nature of mechanical tests is clarified somewhat. Mechanical Experience comes out as a clear factor. Splitting what has been called Spatial Relations into a Visualization and three Space factors needs additional clarification.

Little comfort will be found in this volume for enthusiasts for the Rorschach, TAT, other clinical procedures for diagnosis, and personality inventories (including Bernreuter, Minnesota Multiphasic, and Humm-Wadsworth). Clinical procedures referred to are a clinical interview, observations of behavior during psychomotor testing, observations during a stress situation, observation during a short social situation. None of these procedures was found to be useful. Significant at the one per cent level were only the Thinking Introversion score of Guilford's Inventory with a validity coefficient of −.09 (judged not to be promising), Cooperativeness and Agreeableness of the Guilford-Martin Personnel Inventory, with validity coefficients of .14 and .12. Two categories on the Humm-Wadsworth, Epileptoid and Hysteroid, gave coefficients higher than those in the inventories mentioned above, −.21 and −.18, but they were significant only at the five per cent level since only 202 cases were used. Humm supplied case summaries which did not differentiate passes successfully from failures; however, his summaries were interpreted by the Air Forces' psychologists. Had combat criteria been available, it is possible that some of these personality measures would have been more useful.

---

[2] The reviewer is indebted to Dr. John Dailey for pointing out this fact.

The volume has shortcomings. The mass of material has not been smoothly presented. Clarity is exceeded by bulk. It appears to have been done hastily for the magnitude of the undertaking. The index is incomplete. Research psychometricians will not be seriously bothered by the shortcomings since the book will probably become a bible whose failings will probably come to take on sentimental value in the same way that philatelists treasure a rare stamp with some slight imperfection.

THOMAS W. HARRELL.

University of Illinois.

DAVIS, FREDERICK B. (ED.) *The AAF qualifying examination.* AAF Aviation Psychology Program Research Reports, Report No. 6. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. xvii+266. $1.25.

The AAF Qualifying Examination (originally termed the Aviation Cadet Qualifying Examination) was used as an initial hurdle in the selection of air-crew officers and (after June 1, 1944) in the selection of aerial gunners. The volume under review presents: a history of the examination (Chapter 1); the principles observed in constructing the examination (Chapter 2 and Appendix A); data on the validity, modification, and uses of the examination (Chs. 3 and 4); and studies of specific subtests considered for the examination (Chs. 5–10). Besides the Qualifying Examination, a few other examinations are very briefly reported on, viz.: the Aviation Cadet Educational Examination, the Flight Officer Examination, the English Expression Test, and the Victory Corps Aeronautics Aptitude Test (Chapter 11).

The volume is a meaty one, as evidenced by the 143 numbered tables. No thoughtful reader can plow through this book in an odd afternoon; nor can any conscientious reviewer hope to do it justice within a few pages. In brief overview, what do we learn?

1. Paper-and-pencil tests can be extremely useful in a practical program of selection, even for so "non-academic" a function as flying a warplane. The Qualifying Examination was definitely useful in selecting men who later became successful flying officers. (Chs. 3 and 12.)
2. Research and experimentation led to a significantly improved battery of tests. (Chs. 2 and 3.)
3. Apparatus tests can significantly augment the validity of the paper-and-pencil tests. (Page 39.)
4. Within the limited scope of the reported experiments, paper-and-pencil tests designed to substitute for apparatus tests do not appear to fulfill their intended function. (Ch. 10.)
5. The theoretical possibility of using a "suppression variable" to improve the validity of a battery of tests was in this study carried as far as constructing and administering a special test for the purpose; but unfortunately validity correlations have not yet been obtained. (Ch. 5.)

Some more or less disturbing observations (about which the reviewer feels much less certain that the points above) may be worth mentioning:

1. Correlation coefficients tended to vary more than could be accounted for by random fluctuations of sampling; discouragingly, the variations were frequently in the "wrong" direction (i.e., the validity correlations turned out to be "lower . . . than substantial amounts of data had forecast . . ." (p. 39), or "the promising correlations obtained in earlier samples were not confirmed" (p. 187). A particularly interesting example of excessive variation is pointed out by the author on pp. 191–192; for other instances, see pp. 65, 70 (Table 5.10), 123, 156, and 245–252. The author does not undertake to explain or track down the possible causes of variation.
2. For one reason or another—almost certainly practical, and not due to oversight by the author—certain desirable data were either not collected or not analyzed. Thus, applicants for aviation-cadet training could take the Qualifying Examination repeatedly, "providing only that 30 days elapsed between each testing, and that different forms of the examination were used" (p. 1); yet no data are reported on practice-effects. Similarly, no data are reported on the correlation between successive forms of the Qualifying Examination. Rather often, in the case of a research sample, the answer-sheets containing the necessary basic data became unavailable before statistical analyses could be completed—so that while validity coefficients were computed for a given sample, reliability coefficients and intercorrelations (among the predictor-variables) had to be computed for a substitute-sample of imperfect comparability.[1] Another shortcoming is the rather frequent calculation of validity-data for the *total score* on the battery, but not for the individual *parts* of a battery of tests.

---

[1] Private inquiry indicates that, among various reasons for the unavailability of answer sheets, the principal one was the inaccessibility of data in storage, after transfer of personnel. After such a transfer it was apparently more feasible, on occasion, to collect new data than to obtain the original answer-sheets.

3. A capital lack, in the reviewer's judgment, is the failure to explain in detail whatever strengths or weaknesses the criterion may have held. (Not even an estimated reliability coefficient is offered.) The criterion generally employed is simply "graduation or elimination from elementary pilot training for reasons associated with proficiency in flying" (p. 2). Such a criterion seems extremely general or non-specific: not even the "academic" aspects of training are separated from the non-academic. Specific quantitative criteria would be more useful, not only for the validation and analysis of tests, but probably also as a means of stimulating better observation and evaluation by pilot-instructors. Granted that practical circumstances prevented the use of anything but a simple pass-fail criterion, some brief acknowledgment and warning of the limitations of the criterion would still have been in order.

4. Although the Qualifying Examination was validated against the *pilot* criterion, this same examination had to be passed also by prospective bombardiers, navigators, and (after June 1, 1944) aerial gunners. The administrative convenience of a single screening test is obvious; on the other hand, it seems likely that the various groups of flying personnel have differentiable aptitudes (see e.g., Table 11.2 on p. 220). (This, of course, is a criticism of an administrative practice, rather than of the present research report.)

The central practical question of this volume is: Was the AAF Qualifying Examination valid? The most explicit answer to this question is given in a study of over 1,000 "applicants for aviation-cadet training" who were entered into beginning pilot-training "regardless of their scores on the Qualifying Examination and the Air Crew Classification Tests" (p. 37). In this unselected sample, the validity of the Qualifying Examination (series AC12) was about .45–.50 (against the criterion of graduation from pilot training) (p. 38).

There can be little doubt that the last series of the Qualifying Examination (the AC14 series) was more valid than the first (the AC10 series). Because of the almost inevitable lack of comparability between the figures given for the first and last form, it is not possible to estimate exactly how much improvement in validity was accomplished; the reviewer's tentative judgment is about .20 point of correlation. This appears a remarkably large gain. Part of the gain may conceivably be due to improvement in the reliability or validity of the criterion; the reviewer has no information about this.

In conclusion, the reviewer would like to repeat what was stated at the outset: this book is indeed a solid, meaty work. The exigencies of wartime research have, to be sure, prevented as complete and close knit a series of studies as one might desire. But the author has striven conscientiously to present the information (both favorable and unfavorable) that has been gleaned. The task was immense, and the reviewer cannot close the book without a profound feeling of gratitude for a significant accomplishment. The volume makes contributions both of methods and results.

HERBERT S. CONRAD.

U. S. Office of Education.

GIBSON, JAMES J. (ED.) *Motion picture testing and research.* AAF Aviation Psychology Research Reports, Report No. 7. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. xi+267. $1.50.

This volume is the official report of the Psychological Test Film Unit at Santa Ana Army Air Base, Santa Ana, Calif. Approximately half the report is concerned with the motion picture as a technique and is thus of rather specialized interest. Some of these studies are impressive in the care with which they have been carried out, and some of the results bearing on seating, illumination, and their like are clearly of use to certain readers. It is, however, the other half of the text that will interest most of the readers of the PSYCHOLOGICAL BULLETIN. This portion of the report may be divided into three subdivisions for discussion.

Aptitude tests. The Santa Ana Film Unit was engaged in part in the construction of selection and screening tests for a variety of Air Forces assignments. A number of ingenious motion picture tests for Flying Orientation, Ability to Judge Distance, Estimation of Velocity, Multiple Perception (dial reading) and other abilities were devised and are reported in some detail. The general reader will be struck by the very low correlations between these tests and other tests, between these tests and validating criteria, and even between these tests and themselves (various reliability criteria). Those who were engaged in similar tasks with the armed services will recognize, however, that *r*'s of .25–.55 may often represent genuine achievements. Nevertheless, the low level of predictability makes the permanent usefulness of these studies doubtful, and they must be accepted primarily as guideposts for others who will face such problems in the future.

Airplane recognition research. The investigations which Gibson's group carried out in this area are interesting for two reasons. First, they illustrate the danger of basing large scale activities on inadequate research. The "flash technique" of teaching aircraft recognition was accepted by both Army and Navy for reasons that were supported only by general theoretical considerations. The Santa Ana group has subjected the underlying assumptions of the method to a series of careful tests, only to discover that *none* of those assumptions could be verified in practice. Second, the findings have an intrinsic interest. We now know that training in making shape discriminations at very high speeds does not in itself increase the accuracy of such judgments when longer times are available. We also know that frank analysis of the airplane gestalt into certain component parts that distinguish it from others will aid, at least in the early stages of learning, in making the necessary discrimination.

General perceptual findings. The discussion of investigations related to the perception of aerial space and distance is undoubtedly the most important part of the book from the point of view of the general psychologist. Gibson and Glaser have presented in Chapter 9 a newly organized "system" of explanation for depth perception. It differs in three main ways from traditional systems. First, more emphasis is placed on monocular stimuli. Second, these stimuli are all treated as matters of gradients of retinal stimulation. Third, the treatment of "gradients of velocity" for a moving observer is (so far as the reviewer knows) entirely original. All three of these departures from classical textbook treatments are suggestive and warrant the use of this chapter as supplementary reading for students of visual perception. At least two other contributions deserve mention. One is Gibson's emphasis on position on the vertical meridian as a factor in depth localization. Vertical position (the so-called "stage effect") is an important influence under many conditions of viewing but is often completely omitted from discussions of depth. The other is the quantitative study of size constancy under natural conditions of viewing. The presence of practically complete constancy up to nearly half a mile when only the texture of the terrain gives any distance cues may force a revision of some notions about depth perception based only on laboratory findings and theoretical considerations.

W. C. H. PRENTICE.

Swarthmore College.

MILLER, NEAL E. (ED.) *Psychological research on pilot training.* AAF Aviation Psychology Program Research Reports, Report No. 8. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. xix+488. $1.75.

This volume summarizes research on pilot training accomplished cooperatively by psychologists in the AAF Training Command during World War II. The chief objective of Psychological Research Project (Pilot) was to develop objective measures of flying skill. Allied studies included a job analysis of the primary and specialized functions of student pilots, subjective measures of flying skill, verbal tests of flying information, as well as experiments which evaluate selected training procedures. The book consists of fifteen chapters in addition to six appendices which describe the more technical details of the data. The first fourteen chapters deal mainly with the methods and procedures employed in gathering and analyzing data obtained on large samples of Air Corps personnel under various conditions of flight-training, such as primary, basic and advanced levels, in single-, two-, and multi-engine planes, in contact and instrument flying. An excellent over-all summary of the project, together with recommendations for further work, is to be found in the concluding chapter.

It is scarcely surprising to find, in spite of the unusually skillful and imaginative approach of the investigators that progress in this new and complex area of psychology proved to be hard won. For the most part, coefficients of reliability and validity of the indices ran low, for reasons of a sort already evident to most workers who have attempted psychological research under similar "field conditions." Lack of independent judgments from check-riders who determined the all-important pass-fail criterion, differences in standards from one training school to another (or even within different groups of check-riders at the same school), halo-effects as suggested by strong intercorrelations among different aspects of performance—these are only some of the factors contributing to the variance of the data. Further difficulties inherent in this area of research are emphasized by the finding that pilot performance could be measured with high reliability on the same day even with different observers, but that reliability coefficients tended to drop to zero when the same indices of performance were evaluated on another day by different observers. As the authors emphasize: "The airplane is a complex and variable piece of equipment and the air is an unstable laboratory . . . ."

Among the important accomplishments of this unit was the development of an 81-unit scale for measuring skill in instrument flying, which was sharpened to the point of fairly high reliability for this kind of work (test-retest reliability = .46). More significant still was an experimental study which determined under reasonably well-controlled conditions the extent to which additional training on a low-powered plane transfers to learning the technique of piloting a higher-powered plane. Within the particular circumstances of the training, transfer proved to be small enough to justify the conclusion that "in general 10 weeks of advanced two-engine training, all on the TB-25, was better than 15 weeks of training, 10 of which was on lower-power two-engine planes and only 5 on the TB-25." Again, on the basis of further experimental evidence, it was possible to conclude that, under the conditions of the investigation, "the type of plane in which students receive their basic training does affect their performance in advanced training." Beyond the unique value that such studies have in providing information critical to efficient training, they also demonstrate the feasibility of employing objective measures of flying skill for the quick evaluation of a large scale pilot-training program.

To the reviewer the most important single lesson in the report is the repeated demonstration of the limited value of many "testing" instruments—even when fully exploited, for doing a job that requires basic experimental research in order to obtain conclusive answers. As the authors point out, "While the results of laboratory experiments cannot predict exactly what will happen in a complex situation like that involved in flying training, they can tell one what to look for and how to go about setting up a series of training experiments to decide what kinds and amounts of training . . . will be most efficient in the long run."

J. L. FINAN.

Oberlin College.

KEMP, EDWARD H., AND JOHNSON, A. PEMBERTON. (EDS.) *Psychological research on bombardier training.* AAF Aviation Psychology Program Research Reports, Report No. 9. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. x+294. $1.25.

Perhaps one's first and very general reaction to the military psychological research carried on during the past war is a certain feeling of pride that psychology had come of age and thus was able to make a unique and valuable contribution to the war effort. It is reassuring to learn that many of the concepts, principles, and even equipment which had been developed in the laboratory could be applied so readily to these military problems. To most of the publications within the Army Air Force series the writer reacted in that way.

Because so much money and time of professionally trained personnel was expended in these projects one also expects that they would have done more than apply what we knew, and would have made some contribution to psychological methodology, or to general principles. The contribution of the present study is one of emphasis, an emphasis upon the difficulty in obtaining a reliable criterion.

Superficially it would seem that the problem of obtaining an acceptable criterion for bombardier proficiency should not be difficult. The bombardier's military task is to drop a bomb on a particular target, and this task can be readily duplicated under non-combat conditions. By using a target with concentric rings the magnitude of error could be obtained either by estimations made by the pilot, a fellow bombardier, or by means of aerial photography. One could not quarrel with the validity of such a criterion, for it consisted of a fairly exact duplication of the combat job.

The major difficulty with this criterion, called the circular error, lay in the fact that it had zero reliability and hence could give no precise indication of any particular bombardier's proficiency. Obviously such a criterion had no correlation with any potential predictive measures. A number of studies were conducted to determine the sources of errors in the criterion measure.

In the first place the fault of the measure did not lie in the estimation of the magnitude of error for any single bomb burst. Photographic records could be read with accuracy and the estimates of observers on the magnitude of error were in agreement. A series of experiments indicated that the pilot, the degree of air turbulence, and the particular sight used were all factors contributing to the error. (If precision-built sights differ from each other, what about some of our home-made laboratory equipment?) An extensive experiment which served to control these sources of error was terminated before completion by V-J Day. Unfortunately it has not been continued.

However, the project was not so hampered by the criterion problem that it was unable to make any contribution to bombardier training. It developed tests for selecting bombardier instructors, proficiency tests to measure knowledge of bombing and navigation, and was able to conduct a learning study which gave information on the amount of training necessary to reach a consistent level of performance in bombing accuracy. It was felt by the editors that the project indirectly influenced the bombardier training program through these and other activities so that the quality of training was higher than it otherwise would have been. Their beliefs were supported by a letter of commendation from the Commanding General of the AAF Central Flying Training Command.

DELOS D. WICKENS.

Ohio State University.

CARTER, LAUNOR F. (ED.) *Psychological research on navigator training.* AAF Aviation Psychology Program Research Reports, Report No. 10. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. ix+186. $1.25.

This book describes the research, development and services carried out by the Psychological Research Project (Navigator) under the Army Air Forces Psychology Program during World War II. The following kinds of research problems were conducted by the Project: (1) selection of navigator candidates, (2) navigator training, (3) development of objective ground measures of navigation skill, (4) development of aerial measures of navigation skill, (5) learning studies, (6) studies on instruction in navigator schools, (7) a study of the relative importance of variables involved in air navigation, (8) evaluation of training aids, (9) selection of navigation instructors, and (10) studies of combat conditions, post-combat adjustment, and group therapy. This impressive list of psychologically substantial and militarily useful problems is an indication that it is possible to conduct important research in conjunction with the action program of fighting a war. The navigator selection test battery, for example, had an average validity coefficient of .60 with a pass-fail criterion in advanced navigator school, which is about as well as our present college aptitude tests do in predicting college success.

One of the most interesting jobs described in the book is the development of proficiency tests, both on the ground and in the air, by means of which a more adequate criterion of performance than school grades alone could be obtained. A motion picture test, showing the actual terrain and other navigational information, was produced as part of this program.

The Project evaluated a 48-student training device which simulated on the ground the conditions of navigation in the air. An extensive experiment was designed to discover whether or not the trainer "trained."

As is the case with many training devices possessing high "face validity," the G-2 Trainer proved to have no special efficacy over equivalent time spent in the class room.

The principal new contribution to method in applied psychology is the use of factorial analysis and multiple correlation techniques in scientific job analysis. Error scores on eight and on 14 navigational variables were intercorrelated and two factor analyses were carried out. Both of these analyses identified three important factors. From these studies, it was possible to designate a particular operation (estimating compass deviation) as the primary factor involved in navigator error. This finding led to the recommendation for special training in the operation and suggestions for re-design of the instrument (astro-compass) to minimize the possibility for error. The interrelation of personnel selection, training and instrument design, and the dependence of these activities upon objective, reliable and valid performance measures is beautifully illustrated by these experiments.

The final chapter of the book is an evaluation of this two-year program. Although many problems remain unsolved or partially solved, it is concluded that the selection testing was successful and that the psychologist in the military setting contributed substantially to the solution of many problems not usually considered to be "psychological." Suggestions are also given for the operation of a peacetime research and development department of the Air Forces.

JOHN L. KENNEDY.

Tufts College.

HOBBS, NICHOLAS. (ED.) *Psychological research on flexible gunnery training.* AAF Aviation Psychology Program Research Reports, Report No. 11. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. viii+508. $1.75.

This volume is a review and an evaluation of the work of the Army Air Forces Aviation Psychology Program in the area of flexible gunnery. Although the title of the book implies concern with problems of training only, the research of the flexible gunnery group embraced selection and equipment development problems as well. In fact, the research story told in the volume is convincing in demonstrating the "dog after its own tail" relation between the selection, training, and equipment aspects of psychological research in military or industrial work. The program on flexible gunnery began with a consideration of selection problems and the development of selection tests. The search for adequate criteria, however, against which to validate proposed tests of gunner selection, made it apparent that a thorough investigation of training procedures was necessary. Subsequent effort was directed toward the improvement of training techniques, the development of achievement tests and tests of skill in gunnery maintenance, the evaluation of training devices, and the selection of gunnery instructors. Since these teaching problems were not unrelated to questions of more efficient ways to use available equipment, interest in operating methods developed, leading eventually to inquiries into possible improvements in equipment design through the modification of controls, sights, and other operational features. This elaboration of psychological research interests from selection to equipment problems had many direct parallels in the work of other research groups during the war.

The most important systematic problem discussed by Hobbs and his collaborating authors is that of establishing a satisfactory criterion of gunnery performance. No completely satisfactory criterion was established during the four years of research described in the volume. Combat data, which one might expect to be the best possible kind of criterion, were of limited validity, since records for a particular gunner depended not on his skill alone, but on that of other persons as well, both friend and foe. Various synthetic firing devices designed to train men in the operation of gunnery equipment were of limited "apparent validity" and had reliabilities in the region of .75. Ground firing tests with moving targets had reliabilities between .50 and .70. Air firing at towed targets or with frangible bullets had reliabilities which were generally below .50. In the absence of more satisfactory criteria, therefore, the motion picture record, taken with a gun camera mounted on the gunner's equipment, was accepted as the best check of proficiency. Unfortunately, the reliability of scores obtained from the gun camera films was seriously affected by the type of course flown by the attacking plane and by other temporary conditions. But for both training and criterion purposes, the gun camera was not surpassed by any other type of scoring device.

A concluding evaluation of the general significance of the gunnery research program is provided by the editor:

> The full development . . . of a valuable performance testing technique might be considered important. The studies done on training devices are possibly distinctive by virtue of their comprehensiveness, but the devices themselves have such limited use that this work may best be considered for its wartime applications. Some of the studies on the use of sights may provide suggestions for further research on fundamental problems of visual perception. The training programs developed and evaluated may have long range value as illustrations of the effective application of principles which have long been recognized as sound. Nothing new has been added to knowledge of selection procedures, an adequate job simply having been done to meet an immediate need. Thus, it seems accurate to say that few of these specific findings of gunnery research are of enduring importance.

W. E. KAPPAUF, JR.

Princeton University.

COOK, STUART W. (ED.) *Psychological research on radar observer training.* AAF Aviation Psychology Program Research Reports, Report No. 12. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. iv+340. $1.75.

Research on the selection and training of radar operators was carried out by three groups, the NDRC Project, Sc-70, NS-146, the Aircrew Evaluation and Research Detachment No. 1 of the Eighth Air Force, and the Psychological Research Project (Radar) of the AAF Training Command. This book is primarily a summary of the work of the Radar Project.

Apart from having value as a record of wartime psychological contributions, the report will provide information for those interested in the development of selection tests and standardized proficiency tests, since it was in these two fields that the project concentrated its efforts.

Through the sections on proficiency testing, the extreme difficulties of administration are emphasized. The continuous change and development of equipment, the differing emphases and doctrines in different schools, and the lack of training personnel to provide standardized administration of the proficiency tests and checks, particularly those carried out under the complex conditions of actual flight, all worked against the establishment of the desired evaluation program based upon known reliability and validity.

A distinction is made throughout the report between the observable activities and the psychological factors involved in the radar observer's task. This distinction led, at the outset, to a job description, and to a parallel job analysis, the latter employing primarily factors "definitely or tentatively isolated in the Aviation Psychology Program" (p. 24).

The distinction between description and analysis is reflected in a classification of selection tests and, to a less extent, of proficiency tests, into two types: job analogy tests inspired by the job description, and factor tests suggested by the analysis into psychological components. Validation of the selection tests against proficiency measures indicated that both types must be used to get the best estimate of future performance in the training program. The editor of the report predicts that tests based on psychological analysis will gradually supersede those based on job description.

The primary aim of the project was to establish valid selection instruments. The research purpose of the proficiency testing program was regarded as that of providing criteria for validating the selection tests. In addition, performance on training bombing missions was subjected to a critical analysis, with a view to developing criterion measures. As in other selection research projects with similar aims, where the final task was some performance involving the operation of aircraft, it proved to be much simpler to state the aims and general methods of the validation procedure than to complete significant studies. The insurmountable variation in performance contributed by the complex factors in aircraft operation attenuated the reliability of the radar bombing error scores, just as it did in other attempts to obtain criteria of aerial performance.

The complexities of validating selection tests are thoroughly discussed in a penultimate chapter describing the results of two studies employing three batteries of selection tests validated against training course grades, against an air-crew classification "radar stanine," and, in part, against radar bombing error calculated from camera records obtained during aerial training flights. In spite of the besetting difficulties, a number of significant correlations were obtained. Since the results of these studies were available too late for application in the wartime selection program, their greatest value will lie in indicating to the uninitiated the practical problems which will be present in developing and evaluating tests of this sort.

A final chapter summarizes the principal points of the text and in addition discusses several areas in which training research was planned, including the fields of selecting and training instructors, and of organizing the training curriculum.

C. S. BRIDGMAN.

Navy Special Devices Center.

DAILEY, J. T. (ED.) *Psychological research on flight engineer training.* AAF Aviation Psychology Program Research Reports, Report No. 13. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. iii+227. $1.25.

When the Pearl Harbor debacle gave rise to the immediate necessity of replacing heavy bombardment aircraft in the Pacific, the Army Air Forces decided, because civilian aircraft had handled the route successfully, to fly in the replacements from San Francisco. So many of the planes, however, crash landed at sea from fuel shortages that the Army Air Forces called in civilian companies for consultation. From their experience with long overseas flights, Pan American Airways had found it necessary to add a new crew member whose task it was to plan the flight from the standpoint of thorough knowledge of the mechanics of the craft. Fuel consumption, for example, was found to vary tremendously with the power setting, and one of the major duties of the new position, flight engineer, was to compute the optimal power setting.

It took Pan American two and one-half years to train a flight engineer, but the Air Forces in 1942 did not have two and one-half years. The designs of the B-29 had been on the drawing board since the beginning of the war and were to be in production in 1943, so the Air Forces set in a training program immediately at various schools in the southwest. In 1944 some officers of the Psychological Research Unit No. 2 were assigned the task of preliminary investigation of the requirements and special characteristics desirable for B-29 crew members, and in 1945 the Psychological Research Project (Flight Engineer) was activated with specific directives for the selection and training of flight engineers.

The present report is a history (ch. 1) and description of the work of this project. It is notable for its account of work completed (chs. 4 & 5) in the face of what can only be called major difficulties. Consider, for example, the problem of establishing evaluative and selective criteria (ch. 2) for a new and complex position (1) when the student personnel varied from enlisted men with air mechanic experience to officer graduates of the Yale Technical Training Command Engineer's Officer Candiate School with B-29 training from the Boeing Factory School; (2) when the curriculum was under constant revision of load, content, and geographical locus; (3) when there were long delays and sometimes failures in procurement of mock B-29 equipment; and, finally, (4) the fact that urgent need for flight engineers resulted in extremely few failures and wash-backs, thus vitiating even the simple evaluative aid of grading systems. It is no wonder that many of the research criteria proved to be subjective; and the authors point out the need in this respect of long-time research studies. An appendix to the report includes copies of developed check lists, rating scales, and samples of items from the Flight Engineer Proficiency test.

The job analysis of flight engineer (ch. 3) indicated a responsible and key job in the airplane although the position in the command hierarchy was never specifically established. The flight engineer's duties involved the preparation of the flight plan, pre-flight craft inspection, starting procedures, proper operation of the power plant throughout flight, keeping flight log, post-flight inspection, and writing a post-flight summary report for the ground maintenance crew. His flight station was in the forward pressurized compartment and his instrument panel contained 36 dials, and a control stand fitted with 28 toggle-switches and 14 levers. He had primary authority in emergencies which endangered the loss of the plane.

The sixth and final chapter evaluates the research. Compromises between theory and practice resulted in only moderate successes, but valuables ones nevertheless for clearing the beginning ground of what will certainly be important research of the future. There is a straight-faced inclusion among the factors listed as influencing the trend of research: cessation of hostilities. The word "temporary" can almost be heard.

G. RAYMOND STONE.

University of Oklahoma.

WICKERT, FREDERICK. *Psychological research on problems of redistribution.* AAF Aviation Psychology Program Research Reports, Report No. 14. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. vii+298. $1.25.

This book is a competent and interesting account of the activities of certain American aviation psychologists during World War II. It should prove stimulating to academic psychologists and invaluable to applied psychologists concerned with similar or related problems in the armed forces and elsewhere. The volume contains eight chapters representing major research areas in which redistribution psychologists worked, as follows: (1) history and personnel of the redistribution psychological program; (2) the selection of instructors in the personnel distribution command; (3) combat leadership; (4) combat validation studies; (5) psychological studies of anxiety reaction to combat; (6) surveys of experiences of returned personnel; (7) attitudes and preferences of returned personnel; (8) evaluation of the redistribution psychological program.

The discussions and conclusions are generally distinguished by careful attention to the statistical significance of data based on samples whose basic characteristics including size, variability, manner of selection, and other important relevant factors are specified. Ample tabular material presented and referred to in readily apprehended fashion contribute greatly to the content and readability of the volume. Brief comments on the several types of studies will now be made.

Selection of instructors. To facilitate the further utilization of officer pilots, bombardiers and navigators and enlisted gunners who had been returned from combat duty, the development of test batteries was undertaken. They were designed to indicate general and special abilities, preferences, and other personal characteristics thought to be important for determining assignment to duty as: (1) instructors; (2) operating flying personnel in the Zone of Interior; (3) combat personnel overseas; (4) administrative and general ground personnel. Completion and use of such batteries was confined to the selection of instructors. Satisfactory validation against objective criteria of instructor proficiency was not accomplished during the period in which research was possible. Preliminary validation is not discussed in the volume being reviewed.

Combat leadership. Early studies provided descriptive, anecdotal delineation of behavior of successful and unsuccessful air combat leaders which was used to prepare rating scales. The latter were then employed to check the initial observations. Study of the interrelationships among overseas promotions and objective test and biographical data was a third approach. In all cases several independent samples of adequate size and appropriate characteristics were used to cross-check findings. "The successful combat leader (emerged) as an individual who exhibits outstanding proficiency in his specialty both as a flyer and as a tactical expert, shows consideration for his men, is consistent in his job performance, is sincere and self-sacrificing."

Combat validation studies. Due to the unavailability of useful records on combat performance, a data blank was developed for recording information obtained from the returnees themselves. Included among the items were missions flown, promotions while overseas, aircraft destroyed, highest decoration awarded, precombat flying accidents, flying accidents in combat, position flown, and average overseas efficiency rating. The limitations on the use of these criteria are discussed as are various controls applied as a result of experience gained in previous studies within the Aviation Psychology Program of the relationship between aircrew-candidate test scores and combat criteria. Eleven samples, covering two air forces, three test batteries and the positions of bombardier, navigator, first pilot, copilot and fighter pilot were studied. The N's ranged from 85 to 1125. "To determine the degree to which stanines and test scores were associated with performance in combat as judged by the several criteria, each of the samples was successively dichotomized on the basis of each pertinent criterion, and the means of each of the stanines and test scores for the high and low criterion groups were determined. The significance of each of the differences between the corresponding means of the two groups was investigated by use of Fisher's t test." None of the differences reported was statistically significant, although it is stated that " . . . . some of the results suggest the existence of significant relationships. . . . "

Psychological studies of anxiety reaction to combat. "Operational fatigue" or "anxiety reaction" characterized by tension, tachycardia, sweating, insomnia, nightmares, loss of appetite, irritability, inability to concentrate and low motivation was found by psychiatrists to occur in 5 to 10 per cent of flying officers and 15 to 20 per cent of flying enlisted personnel. No relationship between the aircrew candidate tests or the officer-instructor-selection tests, and anxiety reactions to combat could be found. It may be of interest that the Aiming Stress Test or Steadiness under Pressure, designed specifically to measure resistance to verbally induced emotional responses during a psychomotor performance, had no value for predicting anxiety reaction. The questionnaire known as the Personal Inventory consistently showed a biserial correlation of the order of .50 with psychiatric diagnosis. It should be clearly understood that this does not represent prediction of future behavior since the scores on the Personal Inventory were obtained at approximately the same time the psychiatric diagnosis was made though the two results are independent. The device is merely an aid to psychiatric screening.

Surveys of experiences of returned personnel. This chapter contains a large body of detailed data representing the experiences and impressions of several thousand men in respect to fear and courage in aerial combat, causes of mission-failures, and disorientation-while-flying. Limitations of space preclude any attempt to summarize. The findings contained herein should prove of considerable value to those responsible for training and instrument design.

Attitudes and preferences of returned personnel. As in the case of the preceding chapter, this one does not lend itself to ready summarization. One finding is of particular interest. "How the man reacts psychologically to combat determines his attitude (toward further combat service). The numerical count of hours, missions, and the like, does not." And yet, curiously enough, as stated in the previous chapter, "The hope of returning home after accomplishing a predetermined number of missions was claimed as an important factor in the control of fear. Seventy-five percent favored rotation after a definite number of missions, against 17 percent in favor of having each man fly until . . . . he began to show signs of combat strain."

DONALD E. BAIER.

Personnel Research Section, A.G.O.

BIJOU, SIDNEY W. (ED.) *The psychological program in AAF convalescent hospitals.* AAF Aviation Psychology Program Research Reports, Report No. 15. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. viii+256. $1.25.

In September, 1944 the Army Air Forces Convalescent Hospitals were transferred to the Personnel Distribution Command and a Convalescent Branch of the Psychological Division of that Command was established with Captain Sidney W. Bijou as Officer-in-Charge. Psychologists in the Convalescent Hospitals had five principal functions: convalescent orientation, initial evaluation, progress evaluation, counseling, and research on the psychological problems associated with convalescent activities. At the peak of the program there were fifty officers and eighty-five enlisted men assigned to the psychological services branches of the twelve hospitals. Approximately one hundred and fifty thousand individual services were rendered to approximately thirty thousand convalescent patients. Almost half of the convalescent patients were classified as neuro-psychiatric, with *anxiety reaction* as the most common diagnosis. Typical patients were hospitalized for six to eight weeks exclusive of convalescent leaves and furloughs.

While the psychological programs in the various hospitals differed to some extent, psychologists normally participated first by undertaking a major share of the orientation discussion prior to each patient's transfer from the admitting ward to the convalescent hospital proper. Next, the psychologist conducted an initial evaluation of the patient on the basis of available records, inventories, tests, a case history interview, and a mental status interview. The psychologist then made recommendations to the personal physician, the most important of which had to do with the patient's daily five-hour program of activities. In addition to the services rendered to all incoming patients, psychologists were frequently asked to make special evaluations either to supplement the services of a psychiatrist or to provide personal, vocational, and educational guidance.

Psychological research in the convalescent hospitals was entirely oriented toward producing instruments for more effective services. Some of the instruments, such as the Rorschach, the Shipley Personal Inventory, the Wechsler-Bellevue Scale and the Cornell Selectee Index, were adapted from civilian forms. Others, such as attitude scales specific to the military situation, a biographical inventory and a test entitled Efficiency of Mental Application, were constructed on the spot.

Studies of validity were almost wholly dependent upon psychiatric rather than behavioral criteria and reflect the well-known inadequacies of such criteria. Validity coefficients for the inventory type of instrument under these conditions ran quite high, *e.g.*, a bi-serial *r* of .67 for the revised Shipley Personal Inventory.

Two exploratory studies of biographical information indicated the feasibility of using such items in the statistical prediction of a psychiatrist's estimate of a patient's prognosis. The authors are careful to point out, however, that since the psychiatrists used some of the same biographical information, the criterion is, from one point of view, contaminated.

The Rorschach Test and the Bender-Gestalt Test were studied in relation to the available criteria without conclusive results. This checks with other military studies of this type of instrument. It must be either that Dr. Rorschach's ink spots cast a peculiar magical spell over the psychological mind or that our system of statistical logic urgently needs some extensions. The Incomplete Sentences Test, by way of contrast, demonstrated substantial validity for differentiating groups of patients judged by psychologists to be psychologically unfit for duty ($r=.61$), although its correlation with the psychiatric criterion was considerably smaller ($r=.39$ and .41).

A method for quantifying the inter-personal behavior observed in group counseling was developed. Individual statements by patients and the group leader were classified under such headings as "displacement," "projection," "abreaction," "accepting," "reflecting," etc. Progress during a series of sessions was plotted with respect to the frequencies of the different types of responses.

The appendix contains the formats of several special instruments developed for use in the program.

GEORGE A. KELLY.

Ohio State University.

CRAWFORD, M. P. (ED.) *Psychological research on operational training in the Continental Air Forces.* AAF Aviation Psychology Program Research Reports, Report No. 16. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. vii+367. $1.50.

The avowed purpose of this volume is to "present a record of the research activities of aviation psychologists in the Continental Air Forces"; and " . . . . to provide a summary of psychological research on operational training that will be of value to future research workers in this field." The first of these objectives has been accomplished in an admirable way. The style of the report is concise and clear. The organization is good, and there is little variation in general level of treatment from chapter to chapter—a fact which speaks well for the care with which the editing has been done.

Following an introductory chapter which treats the history, operating procedures, and organization of the Continental Air Forces as these were related to the mission of the research psychologists, there are nine chapters on individual air crew positions. The chapter headings are as follows: Fighter Pilot Studies, Photo-Reconnaissance Pilot, Airplane Commander, Copilot, Navigators, Bombardier, Flight Engineer, Radar Observer, and Flexible Gunnery. In each case, the results of studies designed to determine the relative validities of all classification tests for predicting the success in operational training of the various members of the aircrew are reported and evaluated.

The next three chapters concern the selection and evaluation of lead crews, studies in the acquisition and retention of aircrew skill, and the results of a questionnaire which was designed to determine the attitude toward further duty of aircrew personnel returned from combat. The final chapter is a summary and evaluation. In many ways this is the most valuable chapter in the book, for it is here that the editors present a critical evaluation of the work as a whole.

This reviewer does not intend to be captious when he states that the methods employed, the general type of result obtained, and the difficulties which were encountered are familiar to nearly all psychologists who were engaged in research on military problems. How well the second objective of this volume was accomplished depended in large part on forces over which the authors had no control. If military techniques, training methods and organization change suddenly and radically, the data reported will be of almost no value. On the other hand, the book does serve to acquaint research psychologists with what has already been achieved and need not be repeated. For this purpose, it has been well written.

CARL H. WEDELL.

University of Wisconsin.

LEPLEY, W. M. (ED.) *Psychological research in the theatres of war.* AAF Aviation Psychology Program Research Reports, Report No. 17. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. vi+202. $1.00.

A total of 16 officers and 31 enlisted men engaged in the research activities of the five principal projects with which this report is concerned. The primary missions were two-fold: to obtain first-hand information in the various theaters of war that would serve as combat criterion data against which to validate the original selection and classification tests; and to collect data on the requirements for the various tasks of airmen.

Secondarily, the missions were concerned with the development of proficiency tests for specialized personnel; the evolution of aptitude tests not adequately covered by the original test battery, and such special problems as pilot fatigue in connection with long missions.

As for combat effectiveness, the criterion measures were selected from four categories:

1. Objective measures of combat proficiency, such as accidents, casualties and number of enemy planes shot down.
2. Records of administrative actions taken in combat units (e.g., transfer of inefficient personnel or their reassignment).
3. Ratings based on direct and systematic observation of combat effectiveness in specific missions or analysis of records thereof.
4. Ratings based upon general impressions, reports and incidental observations.

The investigators frankly state the various attenuating and biasing factors that affected the objectivity of their criteria. Included are halo effects in ratings, shrinkage of original populations, goodness of opportunity in determining kill scores of fighter pilots, atmospheric conditions, enemy opposition and deficiencies in record making and keeping, as illustrated by the heterogeneity (both inter-unit and intra-unit) of test batteries employed, scoring formulae applied, the time limits set and modes of recording scores of the selection and classification tests. Other inconveniences and frustrations of a pioneer attempt to fit the best man into each particular job under combat conditions are objectively described.

The report states that "the predictors used in the classification program were, in general, in some degree effective in the selection for combat" (p. 194). Of the 27 predictors for the bombardier category, 18 (66.7%) were predominantly positive. The 37 navigator predictors showed 16 or 43.2% that were positive for the most part. The fighter pilot category indicated 25 (58.1%) of the 43 items as positive as compared with 27 (61.4%) of the 44 predictor items in the bomber pilot category.

The fundamental inadequacy of selection and classification instruments employed is said to reside in their failure to predict the more complex and situational personality characteristics such as "emotional control or stability, judgment, maturity, leadership and motivation" (p. 194)—traits that "experts" judge high among the requirements for successful combat air crew personnel.

N. H. PRONKO.

University of Wichita.

DEEMER, WALTER L. (ED.) *Records, analysis, and test procedures.* AAF Aviation Psychology Program Research Reports, Report No. 18. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. vii+621. $2.25.

This volume is probably the least interesting reading and at the same time the most important of the entire series of AAF Aviation Psychology Research Reports. It is composed of twenty-three chapters divided into three parts and five appendixes. Part I entitled "Test Procedures," ". . . . has a threefold purpose: first to give a brief account of the trial-and-error procedures that characterized early air-crew classification in the AAF Training Command and led to the adoption of standardized testing procedures; second, to present these standardized procedures in detail as an historical record of how standardization was maintained in air-crew classification in World War II; and third, to provide a basis for post-war testing procedures."

Part II, devoted to "Records Procedures," contains an excellent detailed historical description of the record keeping systems employed and should be useful to anyone desiring to use the records for research purposes in the future. One chapter (XIX) gives a detailed description of the procedures for "Obtaining Sums of Squares and Products with the IBM Alphabetical Accounting Machine."

Part III, "Analysis Procedures," is devoted mainly to enumerations and derivations of formulae for statistical analysis of data.

The five appendixes are reprints of various manuals for test administration, classification, and processing of candidates for air-crew. The volume is profusely illustrated with reproductions of record forms, punch cards, and pictures. This volume will be of inestimable value to researchers of the future who may wish to make special studies using data already available because of the minutely detailed descriptions of procedures and methods of record keeping.

E. R. HENRY.

Richardson, Bellows, Henry & Co.

FITTS, P. M. (ED.) *Psychological research on equipment design.* AAF Aviation Psychology Program Research Reports, Report No. 19. Washington, D. C.: U. S. Government Printing Office, 1947. Pp. xii+276. $1.25.

The purpose of this volume is to review the research accomplished on equipment design and related problems of engineering psychology by the Army Air Forces. The work reported on was performed prior to 1 October 1946 at the Department of Psychology, AAF School of Aviation Medicine, and the Psychology Branch, Aero-Medical Laboratory, Engineering Division, Wright Field.

There are 21 chapters all together, 17 of which are reviews of separate research projects written by the individuals who originally carried out the investigations. The first chapter outlines and differentiates the field of engineering psychology, and presents the research areas, methods and techniques and research objectives of this branch of psychology. The second chapter is a survey of sensory display problems in the design of aviation equipment. Chapter 3 is a systematic analysis of human motor abilities related to equipment design and a suggested program of research. The chapters which follow are detailed reports of particular investigations in one or both of these two areas.

Although the chapter reports which constitute the major part of this volume are oriented in terms of the practical problems of equipment design, they contribute, in many instances, to the general body of knowledge in experimental and applied psychology. These research reports may be classified under four different areas as follows:

1. *Visual display problems.* Specific projects are reviewed on the design of clock dials for reading in military time, a comparison of tables versus graphs as a means of presenting numerical data, the design of navigation plotters, the effects of dial diameter and angular separation of scale divisions on dial reading, and a comparison of cartesian-grid versus polar-grid presentation as they affect the interpretability of radar scopes.
2. *Tactual-kinesthetic display problems.* Studies are reported on the ability of pilots to reproduce pressures with the hands or feet, the tactual discrimination of knob shapes, shape coding of control knobs, and the location discrimination ability in the absence of visual cues.
3. *Sensori-motor interrelations.* There are reports on the control of aircraft roll as affected by different attitude indicators, the accuracy of sighting and triggering with different types of gunsight handgrip controls, two projects on the relation between control motions and the movements of visual indicators, the principles of control arrangement for sequential operations, and a comparison of different types of control movements in a compensatory pursuit task.
4. *Environmental effects on visual factors.* For example, efforts were made to study the effects of anoxia on visual illusions and the influence of positive g on the reading of aircraft instrument dials.

Although this book portrays the manner in which experimental psychologists can apply themselves and their methods to military problems, it is apparent that this type of engineering psychology has many civilian and industrial applications. The demonstration that a small change in an indicator or control system can produce a significant improvement in the efficiency of a machine operator is sure to appeal to industrialists. Aside from the practical applications, several novel and important contributions have been made to systematic psychology, particularly the results of studies on kinesthetic and tactual discrimination and the analysis of man's motor abilities. The association of psychologists with engineers has produced some remarkably good experimental apparatus. The manner in which practical military problems have yielded themselves to standard experimental methodology and statistical treatment is a tribute to the ingenuity of the investigators. Inasmuch as each experiment was originally an attack on a specific practical problem, there is occasional difficulty in introducing a particular chapter. On the whole, however, the editor has produced satisfactory continuity and an integrated picture of a new trend in psychology.

LEONARD C. MEAD.

Tufts College.

BOOK REVIEWS

HILGARD, ERNEST R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948. Pp. viii+409. $3.75.

This is a "middle" book. It has the advantages and disadvantages of being in between.

It is half way between the undergraduate and graduate level. Thus it can be used profitably at either level, but must be supplemented at each. Although the writing is clear and simple, only the most advanced undergraduates will have enough background to achieve good understanding without carefully planned supplementary readings and lectures. At the same time, most teachers at the graduate level will want to go beyond this book with assignments from original materials and more detailed discussions. At both levels, the suggested readings at the end of each chapter will be helpful.

This book is also somewhere between a completely impartial survey and a polemic treatise that sets out to make a case against stimulus-response theories and for field-cognitive ones. The reviewer, who hereby warns the reader of his preference for stimulus-response concepts, will have more to say later about the direct and indirect ways in which the author's opinions influence his book.

Finally, it is a middle-sized book. The topics covered and the pages devoted to each are as follows: Introduction, 18; Thorndike, 32; Guthrie, 24; Hull, 40; Skinner, 30; Current Functionalism, 31; Gestalt, 32; Lewin, 25; Wheeler, 27; Tolman, 33; Theories Influenced by Field Conceptions, 31; A Point of View, 37; and References, 34.

In digesting the material in his extensive list of almost 600 references and condensing it to a mere 360 pages of text, the author has performed a difficult task which makes his well-indexed book a convenient and valuable source of titles. On the other hand, a truly rigorous analysis of the logic of learning theories with a critical evaluation of experiments is far too detailed and lengthy a task to be covered completely in the relatively few pages used. Within the wide range of material he considers, the author has had to be highly selective and touch relatively lightly on what he judges to be the high spots.

Quite naturally this necessary selection is influenced by the author's preference for field-cognitive concepts and reaction against stimulus-response ones. For example, he gives approximately 16.5 pages of text to Norman R. F. Maier and only 1.3 to Kenneth Spence. In describing Maier's results he says (p. 307): "These results are taken to be critical of quantitative theories such as those of Spence (1936) which reduce changes in behavior to modifications in the relative strengths of competing tendencies." But Spence's theory and the experiments supporting it are not described. Similarly, with respect to the continuity-discontinuity controversy, the experiments by Lashley and Wade are described in considerable detail (somewhat tempered by a footnote pointing out that all of them have the same unbalanced feature and hence are uncontrolled), while those on the other side (e.g. by McCulloch and Pratt and by Spence) are omitted from the text.

Another type of selection is taking the space to elaborate field-cognitive types of theory beyond their author's published statements, showing how they might fill in gaps and deal with embarassring experimental contradictions, but not doing this for stimulus-response theories. For example, pains are taken to develop what Koffka *might* have said if he had been more systematic about it, while stimulus-response is criticized for not handling punishment adequately; but there is no attempt to develop the relevant S-R concepts, no description of Mowrer's work on anxiety, and a serious misstatement (p. 62) of what Guthrie should predict.

Another way the author's opinion must be expressed is in deciding which concepts to examine critically and which to pass over uncritically. Here the author's critical examination of Thorndike contrasts with his much less critical acceptance of Skinner. Evidence against Thorndike's principles is presented in some detail; such fundamental questions as whether discriminated operants can be shown to follow different laws from conditioned respondents are not raised.

The choice of what material to include, to elaborate, to criticize is a legitimate matter of opinion. Attention has been directed to this only because its influence, though no less real, is much less obvious than that of direct argument. In fact, the introductory student has no way of detecting a bias of this kind.

The author's opinion also shows up occasionally in the choice of emotionally-toned words. For example, the hypothesis of movement-produced stimuli is referred to as an "ever-present alibi" while Wheeler's hypothesis that learning follows the laws of embryology and physical growth is described as an "audacious proposal." Some of the strongest opinions are expressed in informal editorial asides as in the statement (p. 144): "If the notion of operant behavior can be defended and extended, it may do much to break down the rigid stimulus-response conceptions which have too long been dominant in American psychology."

It is the reviewer's opinion that Guthrie does not get a fair hearing. To cite a single example, Guthrie's criticisms of opposing theories are characterized (p. 73): "When he attempts to show why they are not helpful he resorts to superficial criticisms and just occasionally to ridicule." The reviewer agrees that many of Guthrie's criticisms are written in a light and amusing style, but he does not consider them superficial. For example, when Guthrie says, "the rat is left buried in thought," he exposes the central weakness of Tolman's book as well as of many other cognitive approaches, namely, that the theory deals with thought but has no means of getting it over into action. This is a weakness that Tolman has subsequently struggled to correct. An adequate criticism of the chapter on Guthrie is completely beyond the scope of a review; for more details the reader can turn to a discussion by F. D. Sheffield.[1]

Though Hilgard shows his disapproval of the content of stimulus-response theory in a variety of ways, he definitely approves the rigorous logical form of Hull's theory construction and believes that learning science must ultimately be written quantitatively very much as Hull writes it. Though approving their content, Hilgard criticizes the loose form in which field-cognitive theories are expressed. He says (p. 259), "In order for the teachings of gestalt psychology to be assimilated to the ongoing body of scientific knowledge about learning, someone needs to do for gestalt psychology what Hull has been doing for behaviorism, that is, to state the laws or postulates more clearly and precisely, and then to design experiments which will true up these laws for the purposes of scientific prediction in terms of the detailed circumstances under which learning occurs. The suggestions by the gestalt psychologists have been sufficiently fertile that the carrying out of such a program of extension and validation holds rich promise."

It is interesting to compare this suggestion with one issued eighteen years before by Professor Hull: "A friendly invitation is extended to the Gestalt psychologists, and to such other schools as put forward distinctive theories of learning, to exhibit in similar detail a similar deduction from their own principles."[2]

Let us hope that this invitation will not have to be repeated after another eighteen years. When the two theories are put into equally good form, it will be much easier to judge their content.

This book is more than a superficial parade of theories. The careful reader will find that Hilgard has given a great deal of thought to many problems and will find much good common sense with which to agree. Hilgard makes an excellent case for the desirability of attempting to construct rigorous miniature systems and gives a clear description of the way several such systems may eventually be integrated to form a more comprehensive theory. He points out the usefulness of naturalistic observation in providing a broad foundation for theory and experiment, the importance of studying organization and perception, the higher mental processes and the socially-acquired hierarchy of human motives, the need for more attention to the problem of learning capacities, the advantages of young children as experimental subjects, and the importance of selecting the appropriate organism in studying neurological changes correlated with learning.

---

[1] Hilgard's Critique of Guthrie. *Psychol. Rev.* (In press.)

[2] *Psychol. Rev.*, 1930, 37, p. 253.

His own theory of "the provisional try" is stated so sketchily and so modestly subsumed under "A Point of View" that it is difficult to evaluate. He says (p. 337), "Such an interpretation takes all the mystery out of the backward action of effect." The reviewer believes that it substitutes some equally knotty problems. Before any predictions can be made on more than an intuitive basis (i.e. what would I as a reasonable being do if I were in that situation?), the author will have to state the laws governing such things as how a provisional behavior route is kept in suspension, how it has its provisional status changed by confirmation as an appropriate path of action and non-confirmation as an inappropriate one, how the memorial representation of what was done leads to action, and what determines the specific type of action that will be considered appropriate. Perhaps the author will attempt to attack these problems and formulate his own theory in more rigorous detail in a subsequent book.

NEAL E. MILLER.

Yale University.

WOODWORTH, R. S. *Contemporary schools of psychology.* (Rev. Ed.) New York: Ronald Press, 1948. Pp. ix+279. $3.00.

In this new edition, Woodworth is still Woodworth: in style, direct and informal, sometimes even spritely; in substance, erudite, sane, critical but appreciative. Although no new schools have arisen since the first edition in 1931 (thank God, say all of us!) there have been important new developments in the existing schools, and these have led Woodworth to a rather drastic reorganization of his book. Indeed very few passages seem to have survived intact; and there is much fresh material and a fresh appraisal.

Both Structuralism and the Chicago variety of Functionalism, now moribund as distinct schools, receive reduced exposition, though Woodworth calls attention to their permanently valuable contributions. But he now distinguishes a more broadly conceived variety of functionalism with roots going back to the Greeks. (Titchener said it was merely commonsense, which for him meant that, though it might be valuable, it did not have the value of science.) This Functionalism starts "with the results accomplished by the individual and asks how they were accomplished" (p. 32). This is in fact the psychology of "Associationists, Old and New" including Thorndike and the successors of the Chicago Functionalists, of the later Behaviorists, and of nearly all applied psychology. Indeed the psychology of performance, as Woodworth elsewhere calls it, provides a broad platform upon which virtually all psychologists today can stand; and although he is too modest to make such a claim, he is one of the chief architects and builders of this platform.

Woodworth probably underestimates the historical significance of early Behaviorism and the Meyer-Watson-Weiss revolution. Eclectic *au fond du coeur* that he is, Woodworth is not the man to appreciate a merely destructive force, which is just what he shows Behaviorism to have been. Almost nothing constructive, and at the same time distinctive, from that earlier period has survived, nothing, that is, that was not explicit in Functionalism. But it was the revolutionary ardor of Behaviorism which convicted the inheritors of Wundt of sterility; no introspectionist was ever quite the same after reading Watson.

The new Associationists and the new Behaviorists have developed theories rather than schools—Skinner's valiant positivism possibly excepted. Woodworth makes their place in the psychological scheme of things clear; but the graduate student preparing for his examinations must be prepared to go further.

Gestalt, Woodworth rightly sees as a reform of Structuralism; but it has steadily gravitated towards that "primary Functionalism" which is the common denominator of the contemporary scene. Concurrently it is winning almost universal acceptance, if not of Gestalt as a movement, at least of its principal concepts. Woodworth is at home with Gestalt, and most Gestalters will recognize that his treatment is a good introduction.

The analytical psychologists are less likely to be satisfied. Woodworth tries hard to extract the precious metal from the ore of psychoanalysis, but I think he is constitutionally unsympathetic. His chapter is therefore inadequate as an introduction to psychoanalysis. It is, on the other hand, a good statement of what the Freudian movement has so far contributed to psychology.

In his time, McDougall managed to knock together the heads of nearly all his opponents—that is to say, of nearly all other psychologists. "Such blandishments," Woodworth drily remarks, "failed to win votes." But McDougall's contributions to psychology cannot be judged from his manners in controversy. Woodworth shows that his influence and accomplishments were great; the very young men who have been taught by those who remember the twenties to scorn even the mention of McDougall are none the less more and more making motivation a foundation concept (McDougall spoke of purpose or horme), are more and more accepting the notion that motivation cannot, in G. Murphy's phrase, be pumped into an organism by social pressure as air is pumped into a football (McDougall insisted that the motives had an instinctive origin; as Woodworth points out, he does *not* say that our adult motives are instincts—that role he assigned to the sentiments). The psychology of the immediate future will not be McDougallian, yet it seems likely to resemble his doctrine more than that of any of his contemporaries except perhaps Freud.

Holism, formerly treated as part of Gestalt, now shares a chapter with Self-psychology. Again, I do not feel that Woodworth is sympathetic; but if he does not see in these schools great values, neither does he see them as absurd. Instead, as always, he asks what corrective to our thinking is here, what do these intelligent men see that the rest of us are overlooking.

Thus the whole book is an exposition of that much misunderstood position, eclecticism. It is not mere syncretism, not a mere collection of a little that is good from this system or that. Eclecticism is itself a system, and the most intellectually arduous of all. In most systems the ego drive of the proponent is a part of the theory. The system-maker asks of any fact or concept, "Will it enhance my theory?" The eclectic asks, "Will it enrich my vision of truth?" Other systematists ask, "How can I get people to accept my views?" The eclectic asks, "What correction in my views must I make in the light of this evidence, these considerations?" The eclectic walks down "the middle of the road" because that is the best place from which to see what is going on every side.

This book may well be Woodworth's last testament; from it we may all catch a little glimpse of why Aristotle spoke of the truly great man as magnanimous—and of what he meant thereby.

HORACE B. ENGLISH.

Ohio State University.

INDEX OF SUBJECTS

Analysis of experimental data, numerical transformations in, 198
Animal hypnosis, 151
Army Air Forces Aviation Psychology Program Research Reports, special reviews of, 499

Care, infant, and personality, 1
Child psychology, historical beginnings of, 224
Chi-square test, use and misuse of, 433
Colors, Prevost-Fechner-Benham subjective, 97
Comparisons of monaural and binaural thresholds, statistical bias in, 278
Conditioned responses, stimulus generalization of, 337
Control group design, an extension of, 137

Design and psychological experiments, note on Grant's discussion of, 49
Design, extension of control group, 137

Ellis and Conrad, Notes on the "Validity of personality inventories in military practice" by, 303
Extension of control group design, 137

Generalization, stimulus, of conditioned responses, 337
Genetics of schizophrenia, 285
Grant's discussion of the Latin square principle, note on, 49

Historical beginnings of child psychology, 224
Historical survey of psychology in the Soviet Union, 241
Humms' notes, reply to, 307
Hypnosis, animal, 151

Infant care and personality, 1

Latin square principle, note on Grant's discussion of, 49
Learning theory, psychotherapy as a problem in, 366

Misuse of the Chi-square test, use and, 433
Monaural and binaural thresholds, comparisons of, 278

Note on Grant's discussion of the Latin square principle in the design and analysis of psychological experiments, 49
Note on the Postman-Stone controversy, 51
Notes on "The validity of personality inventories in military practice" by Ellis and Conrad, 303
Numerical transformations in the analysis of experimental data, 198

Personality, infant care and, 1
Phenyl-thio-carbamide and related compounds, taste blindness to, 490
Postman-Stone controversy, note on, 51
Prevost-Fechner-Benham subjective colors, 97
Problems of psychometric scatter analysis, 177
Psychometric scatter analysis, problems of, 177
Psychotherapy as a problem in learning theory, 366

Reply to the Humms' "Notes on the 'Validity of personality inventories in military practice,'" 307
Rorschach scores, statistical methods applied to, 393

Scatter analysis, problems of psychometric, 177
Schizophrenia, genetics of, 285
Soviet Union, historical survey of psychology in, 241
Statistical bias in comparisons of monaural and binaural thresholds: binaural summation or binaural supplementation, 278
Statistical methods applied to Rorschach scores, a review, 393

Stimulus generalization of conditioned responses, 337
Subjective colors, Prevost-Fechner-Benham, 97

Transformations in the analysis of data, numerical, 198

Taste blindness to phenyl-thio-carbamide and related compounds, 490

Use and misuse of the Chi-square test, 433

"Validity of personality inventories in military practice," notes on, 303

INDEX OF AUTHORS

ORIGINAL CONTRIBUTIONS, SHORT ARTICLES, SPECIAL REVIEWS, REPORTS, NOTES

Bernard, J., 51
Bugelski, B. R., 49
Burke, C. J., 433
Cohen, J. 97, 490
Conrad, H. S., 307
Cronbach, L. J., 393
Dennis, W., 224
Ellis, A., 307
Gilman, T., 151
Gordon, D. A., 97
Humm, D. G., 303
Humm, K. A., 303
Jastak, J., 177
Lewis, D., 433
Licklider, J. C. R., 278
London, I. D., 241
Marcuse, F. L., 151
Mueller, C. G., 198
Ogdon, D. P., 490
Orlansky, H., 1
Pastore, N., 285
Razran, G., 337
Shoben, E. J., Jr., 366
Smith, M. H., 278
Solomon, R. L., 137

BOOKS REVIEWED

Alschuler, R. H., 327
Amatruda, C., 325
Bayley, N., 317
Berg, I. A., 330
Bijou, S. W., 523
Blum, M. L., 311
Boring, E. G., 54
Bray, C. W., 318
Brown, C. W., 68
Burtt, H. E., 70
Carter, L. F., 515
Cook, S. W., 518
Cottrell, L. S., Jr., 79
Crawford, M. P., 524
Dailey, J. T., 519
Davis, F. B., 508
Davis, R. A., 67
Deemer, W. L., 526
Dennis, W., 57
Doob, L. W., 169
DuBois, P. H., 501
Eberhart, S., 79
Evans, R. M., 166
Ewen, J. H., 323
Eysenck, H. J., 319
Fitts, P. M., 527
Flanagan, J. C., 499
Gesell, A., 325
Ghiselli, E. E., 68
Gibson, J. J., 511
Gorer, G., 89
Guilford, J. P., 506
Hall, V. E., 82
Halstead, W. C., 173
Hattwick, L. B. W., 327
Henle, M., 322
Hilgard, E. R., 529
Hobbs, N., 516
Hollingsworth, H. L., 313
Hsu, H. S., 430
Johnson, A. P., 514
Kaplen, O. J., 74
Kemp, E. H., 514
Langfeld, H. S., 54
Lepley, W. M., 525
Leukart, R. H., 73
Lewis, D., 63
Lindner, R. M., 78
MacKinnon, D. W., 322
Marzolf, S. S., 317
Melton, A. W., 504
Mendel, A. O., 324
Miller, N. E., 512
Moore, T. V., 83
Naumberg, M., 76
Otis, J. L., 73

Pennington, L. A., 330
Peterson, H. A., 317
Poffenberger, A. T., 85
Sechenov, I., 309
Seliger, R. V., 78
Sherif, M., 172
Thorndike, R. L., 503
Tyler, L. E., 326
Vaughan, W. F., 314
Weld, H. P., 54
Wickert, F., 521
Wiener, N., 236
White, R. W., 321
Wolberg, L. R., 81
Woodworth, R. S., 532

BOOK REVIEWERS

Baier, D. E., 521
Bellows, R. M., 501
Black, P., 89
Blair, G. M., 317
Boring, E. G., 57, 309
Bridgman, C. S., 518
Britt, S. H., 79
Bruner, J. S., 169
Conrad, H. S., 508
Corsini, R. J., 78
Davis, F. B., 67
Dulsky, S. G., 74
English, H. B. 532
Finan, J. L., 512
Finch, G., 318
Foley, J. P., Jr., 68
Gardner, J. W., 430
Graham, C. H., 63
Harrell, T. W., 506
Hartmann, G. W., 83, 313
Heathers, G., 172
Henry, E. R., 526
Hunter, W. S., 54
Immergluck, L., 323
Kappauf, W. E., Jr., 516
Kelly, G. A., 523
Kendler, H. H., 322
Kennedy, J. L., 515
Kerr, W. A., 311
Kuenne, M. R., 325
Lanier, L. H., 499
Lawshe, C. H., 73
Lindsley, D. B., 173
Littman, R. A., 314
Lundberg, D. E., 66
Marcuse, F. L., 81
May, M. A., 85
Mead, L. C., 527
Miller, N. E., 529
Mueller, C. G., 63
Norman, R. D., 330
Prentice, W. H. C., 511
Pronko, N. H., 525
Riess, B. F., 82
Richards, T. W., 321
Rickers-Ovsiankina, M. A., 76
Seashore, R. H., 504
Stone, G. R., 519
Stone, L. J., 327
Taylor, F. V., 236
Thompson, A. S., 503
Thorndike, R. L., 326
Tyler, L. E., 319
Wallace, S. R., 70
Wedell, C. H., 524
Wickens, D. D., 514
Wolff, W., 324